

# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/JP05/012631

International filing date: 01 July 2005 (01.07.2005)

Document type: Certified copy of priority document

Document details: Country/Office: JP  
Number: 2004-197296  
Filing date: 02 July 2004 (02.07.2004)

Date of receipt at the International Bureau: 22 July 2005 (22.07.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

日本国特許庁  
JAPAN PATENT OFFICE

01.7.2005

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application: 2004年 7月 2日

出願番号  
Application Number: 特願2004-197296

パリ条約による外国への出願  
に用いる優先権の主張の基礎  
となる出願の国コードと出願  
番号

The country code and number  
of your priority application,  
to be used for filing abroad  
under the Paris Convention, is

JP2004-197296

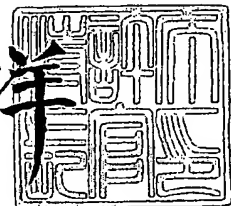
出願人  
Applicant(s):

日本電気株式会社  
エヌイーシーコンピュータテクノ株式会社

2005年 6月 9日

特許庁長官  
Commissioner,  
Japan Patent Office

小川 洋



【書類名】 特許願  
【整理番号】 33510057  
【提出日】 平成16年 7月 2日  
【あて先】 特許庁長官 殿  
【国際特許分類】 G06F 13/00  
G06F 15/00

【発明者】  
【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
【氏名】 細見 岳生

【発明者】  
【住所又は居所】 山梨県甲府市大津町 1088-3 エヌイーシーコンピュータテ  
クノ株式会社内  
【氏名】 渡辺 佳晃

【特許出願人】  
【識別番号】 000004237  
【氏名又は名称】 日本電気株式会社

【特許出願人】  
【識別番号】 000168285  
【氏名又は名称】 エヌイーシーコンピュータテクノ株式会社

【代理人】  
【識別番号】 100102864  
【弁理士】  
【氏名又は名称】 工藤 実

【手数料の表示】  
【予納台帳番号】 053213  
【納付金額】 16,000円

【提出物件の目録】  
【物件名】 特許請求の範囲 1  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1  
【包括委任状番号】 9715177  
【包括委任状番号】 0016289

## 【書類名】特許請求の範囲

## 【請求項1】

ネットワークに接続された複数のプロセッサノードと、  
前記ネットワークに接続された複数の入出力ノードと、  
複数の入出力コントローラとを具備し、  
前記複数のプロセッサノードの各々には、  
複数のプロセッサと、  
複数のデータを格納する主記憶部と、  
前記複数のデータの各々に対してアクセス要求を受け付けることが可能なフリー状態情報が格納されたディレクトリと、  
前記複数のプロセッサと前記主記憶部と前記ディレクトリとに接続されたメモリコントローラとが設けられ、

前記複数の入出力ノードの各々には、

ライトメッセージを発行する複数の入出力デバイスが設けられ、

前記複数の入出力ノードのうちの第1入出力ノードの前記複数の入出力デバイスによって1番目からM番目(Mは1以上の整数である)までのM個のデータに対するM個のライトメッセージが発行されたとき、前記複数の入出力コントローラのうちの第1入出力コントローラは、前記M個のそれぞれのデータに対するM個のライトランザクションを開始し、前記M個のデータのうちの第Iデータ(Iは、 $I=1, 2, \dots, M$ を満たす整数の何れか)は、前記複数のプロセッサノードのうちの第1プロセッサノードをホームとするデータであり、第Iライトメッセージは前記第1プロセッサノードの前記主記憶部に格納された前記複数のデータのうちの第Iデータの値を第Iライトメッセージで指定される値に更新するための命令であり、前記第1入出力コントローラは、前記第Iライトランザクションの処理として、第I書き込み要求メッセージを前記第1プロセッサノードに前記ネットワークを介して出力し、

前記第1プロセッサノードの前記メモリコントローラは、前記第I書き込み要求メッセージを受け取ったとき、前記第Iデータに対して、前記フリー状態情報に代えて、前記第Iデータに対するプロセッサや入出力デバイスからの読み出し要求や他の書き込み要求メッセージを受け付けることができないライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、前記第I書き込み要求メッセージに対して第I書き込み許可メッセージを前記第1入出力コントローラに前記ネットワークを介して出力し、

前記第1入出力コントローラは、

前記第I書き込み許可メッセージを受け取ったときに第Iライトランザクションの更新メッセージ発行処理を行い、

前記第1入出力コントローラは、前記更新メッセージ発行処理において、

第1から第I書き込み許可ランザクションまでのI個の書き込み許可メッセージを既に受け取っているか否かを検査し、

前記I個の書き込み許可メッセージをまだ受け取っていないければ、第I更新メッセージ発行処理を終了し、

前記I個の書き込み許可メッセージを既に受け取っているとき、

前記第Iライトメッセージで指定される値を含む第I更新メッセージを前記第1プロセッサノードに前記ネットワークを介して出力して第Iライトランザクションを完了させ、(I+1)がM以下であれば第(I+1)ライトランザクションの更新メッセージ発行処理を行い、(I+1)がMより大きければ前記第I更新メッセージ発行処理を終了し、

前記第1プロセッサノードの前記メモリコントローラは、前記第I更新メッセージを受け取ったとき、前記第Iデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第1プロセッサノードの前記ディレクトリに格納すると共に、前記第1プロセッサノードの前記主記憶部に格納された前記第Iデータの値を前記第I更新メッセージで指定される値に更新する

マルチプロセッサシステム。

【請求項 2】

請求項 1 に記載のマルチプロセッサシステムにおいて、  
前記複数の入出力ノードには、それぞれ、前記複数の入出力コントローラが更に設けられ、

前記複数の入出力ノードの各々の入出力コントローラ (50-j) には、  
前記複数の入出力ノードの各々の前記複数の入出力デバイスによって発行される前記 M 個のライトメッセージを調停するセクタが更に設けられている

マルチプロセッサシステム。

【請求項 3】

請求項 1 に記載のマルチプロセッサシステムにおいて、

前記ネットワークには、前記複数の入出力コントローラが更に接続され、

前記複数の入出力ノードの各々には、

前記複数の入出力ノードの各々の前記複数の入出力デバイスによって発行される前記 M 個のライトメッセージを調停するセクタが更に設けられている

マルチプロセッサシステム。

【請求項 4】

請求項 1 に記載のマルチプロセッサシステムにおいて、

前記複数のプロセッサノードには、それぞれ、前記複数の入出力コントローラが更に設けられ、

前記複数の入出力ノードの各々には、

前記複数の入出力ノードの各々の前記複数の入出力デバイスによって発行される前記 M 個のライトメッセージを調停するセクタが更に設けられている

マルチプロセッサシステム。

【請求項 5】

請求項 1 に記載のマルチプロセッサシステムにおいて、

前記複数のプロセッサノードと前記複数の入出力ノードとは、それぞれ複数のノードを構成し、

前記複数のノードの各々には、

前記複数のノードの各々の前記複数の入出力デバイスによって発行される前記 M 個のライトメッセージを調停するセクタが更に設けられている

マルチプロセッサシステム。

【請求項 6】

請求項 1 ～ 5 のいずれかに記載のマルチプロセッサシステムにおいて、

前記第 1 プロセッサノードの前記メモリコントローラは、前記第 I データに対して前記ライトロック状態情報が前記第 1 プロセッサノードの前記ディレクトリに格納されているときに、前記第 I データに対する前記第 I 書き込み要求メッセージを前記第 1 入出力コントローラから受け取った場合、前記第 I 書き込み要求メッセージに対して第 I 開放要求メッセージを前記第 1 入出力コントローラに前記ネットワークを介して出力し、

前記第 1 入出力コントローラは、前記第 I 開放要求メッセージを受けて前記第 I 書き込み要求メッセージを前記第 1 プロセッサノードの前記メモリコントローラに前記ネットワークを介して出力すると共に、開放処理を行い、

前記第 1 入出力コントローラは、前記開放処理において、第 K ライトトランザクション {K は、 $K = I + 1, I + 2, \dots, M$  を満たす整数であり、 $I + 1$  は、 $I < (I + 1) < M$  を満たす整数であり、 $I + 2$  は、 $(I + 1) < (I + 2) < M$  を満たす整数である} の進捗を検査し、

未だ第 K 書き込み要求メッセージを発行していない場合、第 K 書き込み要求メッセージの発行を停止し、

既に第 K 書き込み要求メッセージを発行し第 K 書き込み許可メッセージを受け取っている場合、第 K 開放メッセージを前記第 K データのホームである第 2 プロセッサノードに前

記ネットワークを介して出力し、

既に第K書き込み要求メッセージを発行しまだ第K書きこみ許可メッセージを受け取っていない場合は、第K書き込み許可メッセージを受け取った時点で前記第K開放メッセージの発行を行い、

前記第2プロセッサノードの前記メモリコントローラは、前記第K開放メッセージを受け取ったとき、前記第Kデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第2プロセッサノードの前記ディレクトリに格納する

マルチプロセッサシステム。

【請求項7】

請求項1～5のいずれかに記載のマルチプロセッサシステムにおいて、

前記ディレクトリに格納される状態情報には、さらに割り込み可ライトロック状態情報とリクエストロック状態情報を含み、

前記第1入出力コントローラが第I書き込み要求メッセージを発行する際に、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行し、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報がフリー状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラは、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の強・弱書き込み要求メッセージ両方を受け付けることができないライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、第I書き込み許可メッセージを前記第1入出力コントローラに前記ネットワークを介して出力し、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報がフリー状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラは、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができない割り込み可ライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、第I書き込み許可メッセージを前記第1入出力コントローラに前記ネットワークを介して出力し、

前記割り込み可ライトロック状態情報は、この状態にした弱書き込み要求メッセージを発行した入出力コントローラを特定する情報を含み、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報が前記ライトロック状態情報あるいはリクエストロック状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラは、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに前記ネットワークを介して出力し、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報が前記ライトロック状態に格納されている前記第Iデータの状態情報が前記割り込み可ライトロック状態情報、前記ライトロック状態情報、あるいはリクエストロック状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラは、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに前記ネットワークを介して出力し、

前記第1入出力コントローラは、前記第I不許可メッセージを受け取ったとき、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行し、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態

情報が前記割り込み可ライトロック状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラは、前記第I書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記割り込み可ライトロック状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができないリクエストロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、第I再試行要求メッセージを前記ディレクトリに格納された入出力コントローラを特定する情報が指す第2入出力コントローラに宛てて前記ネットワークを介して出力し、

前記第I再試行要求メッセージを受け取った前記第2入出力コントローラは、再試行処理を行う

マルチプロセッサシステム。

【請求項8】

請求項7に記載のマルチプロセッサシステムにおいて、

前記第2入出力コントローラが行う再試行処理は、

第I開放メッセージを前記第1プロセッサノードの前記メモリコントローラに前記ネットワークを介して出力し、

前記第Iデータとアドレスを同じくするライトランザクションで書き込み許可メッセージを受け取り済みのものに関して、まだ更新メッセージを発行していなければ更新メッセージの発行を停止して、前記第I開放メッセージの発行後に、前記第1プロセッサノードに宛てて書き込み要求メッセージを発行し、

前記第I開放メッセージを受けた前記第1プロセッサノードの前記メモリコントローラは、前記リクエストロック状態情報に代えてライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、第I書き込み許可メッセージを前記第1入出力コントローラに宛てて前記ネットワークを介して出力する

マルチプロセッサシステム。

【請求項9】

複数のプロセッサノードと、複数の入出力ノードと、複数の入出力コントローラとを具備し、

前記複数のプロセッサノードの各々には、複数のプロセッサと、複数のデータを格納する主記憶部と、前記複数のデータの各々に対してアクセス要求を受け付けることが可能なフリー状態情報が格納されたディレクトリと、前記複数のプロセッサと前記主記憶部と前記ディレクトリとに接続されたメモリコントローラとが設けられ、

前記複数の入出力ノードの各々には、ライトメッセージを発行する複数の入出力デバイスが設けられた、

マルチプロセッサシステムに適用されるメモリアクセス処理方法であって、

前記メモリアクセス処理方法は、

前記複数の入出力ノードのうちの第1入出力ノードの前記複数の入出力デバイスによって1番目からM番目(Mは1以上の整数である)までのM個のデータに対するM個のライトメッセージが発行されたとき、前記複数の入出力コントローラのうちの第1入出力コントローラが、前記M個のそれぞれのデータに対するM個のライトランザクションを開始し、前記M個のデータのうちの第Iデータ(Iは、 $I=1, 2, \dots, M$ を満たす整数の何れか)は、前記複数のプロセッサノードのうちの第1プロセッサノードをホームとするデータであり、第Iライトメッセージは前記第1プロセッサノードの前記主記憶部に格納された前記複数のデータのうちの第Iデータの値を第Iライトメッセージで指定される値に更新するための命令であり、前記第1入出力コントローラが、前記第Iライトランザクションの処理として、第I書き込み要求メッセージを前記第1プロセッサノードに出力するステップと、

前記第1プロセッサノードの前記メモリコントローラが、前記第I書き込み要求メッセージを受け取ったとき、前記第Iデータに対して、前記フリー状態情報に代えて、前記第Iデータに対するプロセッサや入出力デバイスからの読み出し要求や他の書き込み要求メ

ッセージを受け付けることができないライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、前記第I書き込み要求メッセージに対して第I書き込み許可メッセージを前記第1入出力コントローラに出力するステップと、

前記第1入出力コントローラが、前記第I書き込み許可メッセージを受け取ったとき、第Iライトトランザクションの処理として更新メッセージ発行処理を行うステップとを含み、

前記第1入出力コントローラが行う前記更新メッセージ発行処理は、

第1から第I書き込み許可メッセージまでのI個の書き込み許可メッセージを既に受け取っているか否かを検査するステップと、

前記I個の書き込み許可メッセージをまだ受け取っていないければ、前記第I更新メッセージ発行処理を終了するステップと、

前記I個の書き込み許可メッセージを既に受け取っていれば、前記第Iライトメッセージで指定される値を含む第I更新メッセージを前記第1プロセッサノードに出力して第Iライトトランザクションを完了させ、 $(I+1)$ がM以下であれば第 $(I+1)$ ライトトランザクションの更新メッセージ発行処理を行い、 $(I+1)$ がMより大きければ前記第I更新メッセージ発行処理を終了するステップとを含み、

前記メモリアクセス処理方法は、更に、

前記第1プロセッサノードの前記メモリコントローラが、前記第I更新メッセージを受け取ったとき、前記第Iデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第1プロセッサノードの前記ディレクトリに格納すると共に、前記第1プロセッサノードの前記主記憶部に格納された前記第Iデータの値を前記第I更新メッセージで指定される値に更新するステップと

を含むメモリアクセス処理方法。

#### 【請求項10】

請求項9に記載のメモリアクセス処理方法において、

更に、

前記複数の入出力ノードの各々の前記複数の入出力デバイスによって発行される前記M個のライトメッセージを調停するステップ

を含むメモリアクセス処理方法。

#### 【請求項11】

請求項9又は10に記載のメモリアクセス処理方法において、

更に、

前記第Iデータに対して前記ライトロック状態情報が前記第1プロセッサノードの前記ディレクトリに格納されているときに、前記第1プロセッサノードの前記メモリコントローラが、前記第Iデータに対する前記第I書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記第I書き込み要求メッセージに対して第I開放要求メッセージを前記第1入出力コントローラに出力するステップと、

前記第1入出力コントローラが、前記第I開放要求メッセージを受けて前記第I書き込み要求メッセージを前記第1プロセッサノードの前記メモリコントローラに出力すると共に、開放処理を行なうステップとを含み、

前記開放処理を行なうステップは、

前記第Iデータに後続する第Kライトトランザクション  $\{K$ は、 $K=I+1, I+2, \dots, M$ を満たす整数であり、 $I+1$ は、 $I < (I+1) < M$ を満たす整数であり、 $I+2$ は、 $(I+1) < (I+2) < M$ を満たす整数である $\}$ の進捗を検査し、未だ第K書き込み要求メッセージを発行していない場合、第K書き込み要求メッセージの発行を停止するステップと、

既に第K書き込み要求メッセージを発行し第K書き込み許可メッセージを受け取っている場合、第K開放メッセージを前記第Kデータのホームである第2プロセッサノードの前記メモリコントローラに出力するステップと、

既に第K書き込み要求メッセージを発行しまだ第K書き込み許可メッセージを受け取っ



ていない場合は、第K書き込み許可メッセージを受け取った時点で前記第K開放メッセージの発行を行うステップとを含み、

前記メモリアクセス処理方法は、更に、

前記第2プロセッサノードの前記メモリコントローラが、前記第K開放メッセージを受け取ったとき、前記第Iデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第2プロセッサノードの前記ディレクトリに格納するステップを含むメモリアクセス処理方法。

【請求項12】

請求項9又は10に記載のメモリアクセス処理方法において、

前記ディレクトリに格納される状態情報には、さらに割り込み可ライトロック状態情報とリクエストロック状態情報を含み、

前記メモリアクセス処理方法は、更に、

前記第1入出力コントローラが第I書き込み要求メッセージを発行する際に、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行するステップと、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報がフリー状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラが、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の強・弱書き込み要求メッセージ両方を受け付けることができないライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、第I書き込み許可メッセージを前記第1入出力コントローラに出力するステップと、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報がフリー状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラが、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができない割り込み可ライトロック状態情報を前記第1プロセッサノードの前記ディレクトリに格納し、第I書き込み許可メッセージを前記第1入出力コントローラに出力するステップと、

前記割り込み可ライトロック状態情報は、この状態にした弱書き込み要求メッセージを発行した入出力コントローラを特定する情報を含み、

前記メモリアクセス処理方法は、更に、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報が前記ライトロック状態情報あるいはリクエストロック状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラが、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに出力するステップと、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報が前記ライトロック状態に格納されている前記第Iデータの状態情報が前記割り込み可ライトロック状態情報、前記ライトロック状態情報、あるいはリクエストロック状態情報であるときに、前記第1プロセッサノードの前記メモリコントローラが、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに出力するステップと、

前記第1入出力コントローラが、前記第I不許可メッセージを受け取ったとき、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行するステップと、

前記第1プロセッサノードの前記ディレクトリに格納されている前記第Iデータの状態情報が前記割り込み可ライトロック状態情報であるときに、前記第1プロセッサノードの

前記メモリコントローラが、前記第 I 強書き込み要求メッセージを前記第 1 入出力コントローラから受け取った場合、前記割り込み可ライトロック状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第 I データに対する他の弱書き込み要求メッセージを受け付けることができないリクエストロック状態情報を前記第 1 プロセッサノードの前記ディレクトリに格納し、第 I 再試行要求メッセージを前記ディレクトリに格納された入出力コントローラを特定する情報が指す第 2 入出力コントローラに宛てて出力するステップと、

前記第 I 再試行要求メッセージを受け取った前記第 2 入出力コントローラが、再試行処理を行うステップと

を含むメモリアクセス処理方法。

【請求項 13】

請求項 12 に記載のメモリアクセス処理方法において、

前記第 2 入出力コントローラが再試行処理を行うステップは、

第 I 開放メッセージを前記第 1 プロセッサノードの前記メモリコントローラに出力するステップと、

前記第 I データとアドレスを同じくするライトトランザクションで書き込み許可メッセージを受け取り済みのものに関して、まだ更新メッセージを発行していなければ更新メッセージの発行を停止して、前記第 I 開放メッセージの発行後に、前記第 1 プロセッサノードに宛てて書き込み要求メッセージを発行するステップと、

前記第 I 開放メッセージを受けた前記第 1 プロセッサノードの前記メモリコントローラは、前記リクエストロック状態情報に代えてライトロック状態情報を前記第 1 プロセッサノードの前記ディレクトリに格納し、第 I 書き込み許可メッセージを前記第 1 入出力コントローラに宛てて出力するステップと

を含むメモリアクセス処理方法。

## 【書類名】明細書

【発明の名称】マルチプロセッサシステムおよびメモリアクセス処理方法

## 【技術分野】

## 【0001】

本発明は、メモリを共有する密結合型のマルチプロセッサシステムに関し、入出力デバイスからのメモリアクセスを処理するマルチプロセッサシステムおよびメモリアクセス処理方法に関する。

## 【背景技術】

## 【0002】

PCIバス仕様リビジョン2.1等の規定では、PCIバス上の入出力デバイスを要求元にしたライトメッセージは順序を保障しなければならないという制約がある。即ち、先行のメッセージが完了してから後続のメッセージが完了することを保障しなければならない。

## 【0003】

図1はディレクトリ方式によってデータの一貫性を維持するマルチプロセッサシステムの構成を示す図である。マルチプロセッサシステムは、複数のプロセッサノード101-1~101-m (mは1以上の整数)と、複数の入出力ノード103-1~103-n (nは1以上の整数)とを具備している。複数のプロセッサノード101-1~101-mと複数の入出力ノード103-1~103-nとは、ネットワーク102に接続され、外部からのクロックに応じて動作する。プロセッサノード101-i (i=1, 2, ..., m)には、プロセッサ110-i-1、110-i-2と、ディレクトリ120-iと、主記憶部(メモリ)130-iと、メモリコントローラ140-iとが設けられている。入出力ノード103-j (j=1, 2, ..., n)は、入出力コントローラ150-jと、外部からの命令によりメッセージを発行する複数の入出力デバイス160-j-1、160-j-2とが設けられている。

主記憶部130-iには、複数のデータが格納されている。複数のデータの各々は、その内容を表す値を含んでいる。

## 【0004】

図1に示すようなマルチプロセッサシステムにおいて前記ライトメッセージの順序制約を満たす従来技術を紹介する。特許文献1の「発明が解決しようとする課題」に記載された技術(以降、従来技術1)は、先行するライトメッセージの完了が保障されるまで後続するライトメッセージの発行を留めることでこの制約を満たすというものである。

## 【0005】

図2を参照しながら、順序制約のあるライトメッセージが連続して発行された場合の従来技術1の動作を説明する。図2は、入出力デバイス160-1-1が、データA、B、Cに対するライトメッセージをそれぞれステップ1、2、3で発行した場合の動作を示している。ここで、データAおよびデータBはプロセッサノード1-1をホームとし、データCはプロセッサノード1-2をホームとするデータであるとする。また、1ステップは1クロックに対応する。

## 【0006】

入出力コントローラ150-1は、ステップ2にて、入出力デバイス160-1-1からのライトメッセージとしてライトAメッセージを受け取る。このとき、入出力コントローラ150-1は、ステップ3にて、ライトAメッセージで指定される値を含む更新Aメッセージをホームのメモリコントローラ140-1に宛ててネットワーク102に出力する。

メモリコントローラ140-1は、ステップ4にて、入出力コントローラ150-1からの更新Aメッセージを受けて、主記憶部130-1に格納されたデータの値を更新Aメッセージで指定される値に更新する。このとき、メモリコントローラ140-1は、ステップ5にて、入出力コントローラ150-1に宛ててネットワーク102に完了Aメッセージを出力する。

入出力コントローラ150-1は、ステップ6にて、メモリコントローラ140-1からの完了Aメッセージを受け取り、先行するライトAが完了したことを認識する。

【0007】

入出力コントローラ150-1は、ステップ3にて、第2ライトメッセージとしてライトBメッセージを受け取る。このとき、入出力コントローラ150-1は、先行するライトAメッセージが完了するステップ6までライトBメッセージを留め置き、ステップ7にて、ライトBメッセージで指定される値を含む更新Bメッセージをメモリコントローラ140-1に宛ててネットワーク102に出力する。

メモリコントローラ140-1は、ステップ8にて、入出力コントローラ150-1からの更新Bメッセージを受けて、主記憶部130-1に格納されたデータの値を更新Bメッセージで指定される値に更新する。このとき、メモリコントローラ140-1は、ステップ9にて、入出力コントローラ150-1に宛ててネットワーク102に完了Bメッセージを出力する。

入出力コントローラ150-1は、ステップ10にて、メモリコントローラ140-1からの完了Bメッセージを受け取り、先行するライトBが完了したことを認識する。

【0008】

入出力コントローラ150-1は、ステップ4にて、第3ライトメッセージとしてライトCメッセージを受け取る。このとき、入出力コントローラ150-1は、先行するライトAメッセージおよびライトBメッセージが両方とも完了するステップ10までライトCメッセージを留め置き、ステップ11にて、ライトCメッセージで指定される値を含む更新Cメッセージをメモリコントローラ140-2に宛ててネットワーク102に出力する。

メモリコントローラ140-2は、ステップ12にて、入出力コントローラ150-1からの更新Cメッセージを受けて、主記憶部130-2に格納されたデータの値を更新Cメッセージで指定される値に更新する。このとき、メモリコントローラ140-2は、ステップ13にて、入出力コントローラ150-1に宛ててネットワーク102に完了Cメッセージを出力する。

入出力コントローラ150-1は、ステップ14にて、メモリコントローラ140-2からの完了Cメッセージを受け取り、ライトCが完了したことを認識する。

【0009】

このように、入出力コントローラ150-jは、先行するライトメッセージが完了してから次のライトメッセージを発行することで、入出力デバイス160-j-1、60-j-2が発行した複数のライトメッセージの順序を保障することができる。しかし、3つのライトメッセージを処理するのに14ステップを要する。

【0010】

このライトメッセージを処理するのに要する時間が長く、性能が劣化する問題を解決する従来技術（従来技術2）が特許文献1に記載されている。この技術は、同一プロセッサノードを宛先とするライトメッセージの連続発行、即ち先行するメッセージの完了が保障される前に後続するメッセージを発行することを可能とするものである。

【0011】

図3を参照しながら、従来技術2の動作を説明する。

【0012】

入出力コントローラ150-1は、ステップ2にて、第1ライトメッセージとしてライトAメッセージを受け取り、ライトAメッセージがメモリコントローラ140-1をホームとするライトであることを認識する。このとき、入出力コントローラ150-1は、ステップ3にて、ライトAメッセージで指定される値を含む更新Aメッセージをホームのメモリコントローラ140-1に宛ててネットワーク102に出力する。

メモリコントローラ140-1は、ステップ4にて、入出力コントローラ150-1からの更新Aメッセージを受けて主記憶部130-1のデータを更新する。このとき、メモリコントローラ140-1は、ステップ5にて、入出力コントローラ150-1に宛ててネットワーク102に完了Aメッセージを出力する。

入出力コントローラ150-1は、ステップ6にて、メモリコントローラ140-1からの完了Aメッセージを受け取り、先行するライトAメッセージが完了したことを認識する。

【0013】

入出力コントローラ150-1は、ステップ3にて、第2ライトメッセージとしてライトBメッセージを受け取り、ライトBメッセージがメモリコントローラ140-1をホームとするライトであることを認識し、先行するライトAメッセージと同じホーム（メモリコントローラ140-1）であることを認識する。このとき、入出力コントローラ150-1は、ライトAメッセージの完了を待つことなく、ステップ4にて、ライトBメッセージで指定される値を含む更新Bメッセージをホームのメモリコントローラ140-1に宛ててネットワーク102に出力する。

メモリコントローラ140-1は、ステップ5にて、入出力コントローラ150-1からの更新Bメッセージを受けて主記憶部130-1のデータを更新する。このとき、メモリコントローラ140-1は、ステップ6にて、入出力コントローラ150-1に宛ててネットワーク102に完了Bメッセージを出力する。

入出力コントローラ150-1は、ステップ7にて、メモリコントローラ140-1からのステップ7で完了Bメッセージを受け取る。

【0014】

入出力コントローラ150-1は、ステップ4にて、第3ライトメッセージとしてライトCメッセージを受け取り、ライトCメッセージがメモリコントローラ140-2をホームとするライトであることを認識し、先行するライトAメッセージ、ライトBメッセージとはホーム（メモリコントローラ140-1）が異なることを認識する。このとき、入出力コントローラ150-1は、両ライトメッセージ（ライトAメッセージ、ライトBメッセージ）とも完了するステップ7までライトCメッセージを留め置き、ステップ8にて、ライトCメッセージで指定される値を含む更新Cメッセージをメモリコントローラ140-2に宛ててネットワーク102に出力する。

メモリコントローラ140-2は、ステップ9にて、入出力コントローラ150-1からの更新Cメッセージを受けて主記憶部130-2のデータを更新する。このとき、メモリコントローラ140-2は、ステップ10にて、入出力コントローラ150-1に宛ててネットワーク102に完了Cメッセージを出力する。

入出力コントローラ150-1は、ステップ11にて、メモリコントローラ140-1からの完了Cメッセージを受け取り、ライトCが完了したことを認識する。

【0015】

このように、入出力コントローラ150-jは、ホームを同じにするライトメッセージについては連続して発行し、ホームが異なるライトメッセージについては先行するライトメッセージが完了してから発行する。ネットワーク102は2点間のメッセージの順序を保障するので、上記例ではライトAメッセージとライトBメッセージはその順でメモリコントローラ140-1に到着することが保障される。そのため、ライトAメッセージを追い越してライトBメッセージが先にメモリコントローラ140-1で処理されることはなく順序を保障することができる。

【0016】

しかし、この従来技術2でも3つのライトメッセージを処理するのに11ステップを要する。

【0017】

【特許文献1】特開2001-216259号公報

【発明の開示】

【発明が解決しようとする課題】

【0018】

上記の従来技術1、従来技術2では、入出力コントローラが、異なるプロセッサノードを宛先とする入出力デバイスからの複数のライトメッセージを連続して処理することがで

きない。このため、入出力コントローラがライトメッセージの処理を行う場合に要する時間は、長くなってしまう。

#### 【0019】

本発明の課題は、入出力コントローラが、異なるプロセッサノードを宛先とする入出力デバイスからの複数のライトメッセージを連続して処理することができるマルチプロセッサシステムを提供することにある。

本発明の他の課題は、入出力コントローラがライトメッセージの処理を行う場合に要する時間を短くすることができるマルチプロセッサシステムを提供することにある。

#### 【課題を解決するための手段】

#### 【0020】

以下に、[発明を実施するための最良の形態]で使用する番号・符号を用いて、課題を解決するための手段を説明する。これらの番号・符号は、[特許請求の範囲]の記載と[発明を実施するための最良の形態]の記載との対応関係を明らかにするために付加されたものであるが、[特許請求の範囲]に記載されている発明の技術的範囲の解釈に用いてはならない。

#### 【0021】

本発明のマルチプロセッサシステムは、ネットワーク(2)に接続された複数のプロセッサノード(1-1~1-m) (mは1以上の整数である)と、前記ネットワーク(2)に接続された複数の入出力ノード(3-1~3-n) (nは1以上の整数である)と、複数の入出力コントローラとを具備している。

前記複数のプロセッサノードの各々(1-i) (i=1, 2, ..., m)には、複数のプロセッサ(10-i-1, 10-i-2)と、複数のデータを格納する主記憶部(30-i)と、前記複数のデータの各々に対してアクセス要求を受け付けることが可能なフリー状態情報が格納されたディレクトリ(20-i)と、前記複数のプロセッサ(10-i-1, 10-i-2)と前記主記憶部(30-i)と前記ディレクトリ(20-i)とに接続されたメモリコントローラ(40-i)とが設けられている。

前記複数の入出力ノードの各々(3-j) (j=1, 2, ..., n)には、ライトメッセージを発行する複数の入出力デバイス(60-j-1, 60-j-2)が設けられている。

前記複数の入出力ノード(3-1~3-n)のうちの第1入出力ノード(3-1)の前記複数の入出力デバイス(60-1-1, 60-1-2)によって1番目からM番目(Mは1以上の整数である)までのM個のデータに対するM個のライトメッセージが発行されたとき、前記複数の入出力コントローラのうちの第1入出力コントローラは、前記M個のそれぞれのデータに対するM個のライトランザクションを開始する。前記M個のデータのうちの第Iデータ(Iは、I=1, 2, ..., Mを満たす整数の何れか)は、前記複数のプロセッサノード(1-1~1-m)のうちの第1プロセッサノード(1-i)をホームとするデータである。第Iライトランザクションは前記第1プロセッサノード(1-i)の前記主記憶部(30-i)に格納された前記複数のデータのうちの第Iデータの値を第Iライトランザクションで指定される値に更新するための命令である。前記第1入出力コントローラは、前記第Iライトランザクションの処理として、第I書き込み要求メッセージを前記第1プロセッサノード(1-i)に前記ネットワーク(2)を介して出力する。

前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I書き込み要求メッセージを受け取ったとき、前記第Iデータに対して、前記フリー状態情報に代えて、前記第Iデータに対するプロセッサや入出力デバイスからの読み出し要求や他の書き込み要求メッセージを受け付けることができないライトロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、前記第I書き込み要求メッセージに対して第I書き込み許可メッセージを前記第1入出力コントローラに前記ネットワーク(2)を介して出力する。

前記第I書き込み許可メッセージを受け取ったときに第Iライトランザクションの更

新メッセージ発行処理を行う。前記第1入出力コントローラは、前記更新メッセージ発行処理において、第1から第I書き込み許可トランザクションまでのI個の書き込み許可メッセージを既に受け取っているか否かを検査し、前記I個の書き込み許可メッセージをまだ受け取っていないければ、第I更新メッセージ発行処理を終了し、前記I個の書き込み許可メッセージを既に受け取っているとき、前記第Iライトメッセージで指定される値を含む第I更新トランザクションを前記第1プロセッサノードに前記ネットワーク(2)を介して出力して第Iライトトランザクションを完了させ、(I+1)がM以下であれば第(I+1)ライトトランザクションの更新メッセージ発行処理を行い、(I+1)がMより大きければ前記第I更新メッセージ発行処理を終了する。

前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I更新メッセージを受け取ったとき、前記第Iデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納すると共に、前記第1プロセッサノード(1-i)の前記主記憶部(30-i)に格納された前記第Iデータの値を前記第I更新メッセージで指定される値に更新する。

#### 【0022】

本発明のマルチプロセッサシステムにおいて、前記複数の入出力ノード(3-1~3-n)には、それぞれ、前記複数の入出力コントローラ(50-1~50-n)が更に設けられている。

前記複数の入出力ノードの各々(3-j)(j=1, 2, ..., n)の入出力コントローラ(50-j)には、前記複数の入出力ノードの各々(3-j)の前記複数の入出力デバイス(60-j-1, 60-j-2)によって発行される前記M個のライトメッセージを調停するセクタ(71-j)が更に設けられている。

#### 【0023】

本発明のマルチプロセッサシステムにおいて、前記ネットワーク(2)には、前記複数の入出力コントローラ(52-1~52-n)が更に接続されている。

前記複数の入出力ノードの各々(3-j)(j=1, 2, ..., n)には、前記複数の入出力ノードの各々(3-j)の前記複数の入出力デバイス(60-j-1, 60-j-2)によって発行される前記M個のライトメッセージを調停するセクタ(51-j)が更に設けられている。

#### 【0024】

本発明のマルチプロセッサシステムにおいて、前記複数のプロセッサノード(1-1~1-m)には、それぞれ、前記複数の入出力コントローラ(52-1~52-m)(図示しない)が更に設けられている。

前記複数の入出力ノードの各々(3-j)(j=1, 2, ..., n)には、前記複数の入出力ノードの各々(3-j)の前記複数の入出力デバイス(60-j-1, 60-j-2)によって発行される前記M個のライトメッセージを調停するセクタ(51-j)が更に設けられている。

#### 【0025】

本発明のマルチプロセッサシステムにおいて、前記複数のプロセッサノード(1-1~1-m)と前記複数の入出力ノード(3-1~3-n)とは、それぞれ複数のノード(図示しない)(m=n)を構成する。

前記複数のノード(図示しない)の各々には、前記複数のノード(図示しない)の各々の前記複数の入出力デバイス(60-j-1, 60-j-2)によって発行される前記M個のライトメッセージを調停するセクタ(51-j)が更に設けられている。

#### 【0026】

本発明のマルチプロセッサシステムにおいて、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第Iデータに対して前記ライトロック状態情報が前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されているときに、前記第Iデータに対する前記第I書き込み要求メッセージを前記第1入出

力コントローラから受け取った場合、前記第I書き込み要求メッセージに対して第I開放要求メッセージを前記第1入出力コントローラに前記ネットワーク(2)を介して出力する。

前記第1入出力コントローラは、前記第I開放要求メッセージを受けて前記第I書き込み要求メッセージを前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)に前記ネットワーク(2)を介して出力すると共に、開放処理を行なう。

前記第1入出力コントローラは、前記開放処理において、第Kライトトランザクション{Kは、 $K=I+1, I+2, \dots, M$ を満たす整数であり、 $I+1$ は、 $I < (I+1) < M$ を満たす整数であり、 $I+2$ は、 $(I+1) < (I+2) < M$ を満たす整数である}の進捗を検査し、未だ第K書き込み要求メッセージを発行していない場合、第K書き込み要求メッセージの発行を停止する。既に第K書き込み要求トランザクションを発行し第K書き込み許可トランザクションを受け取っている場合、前記第1入出力コントローラは、第K開放トランザクションを前記第Kデータのホームである第2プロセッサノード(1-k)( $k=1, 2, \dots, m$ )に前記ネットワーク(2)を介して出力する。既に第K書き込み要求メッセージを発行しまだ第K書き込み許可メッセージを受け取っていない場合は、前記第1入出力コントローラは、第K書き込み許可メッセージを受け取った時点で前記第K開放メッセージの発行を行う。

前記第2プロセッサノード(1-k)の前記メモリコントローラ(40-k)は、前記第K開放メッセージを受け取ったとき、前記第Kデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第2プロセッサノード(1-k)の前記ディレクトリ(20-k)に格納する。

#### 【0027】

前記ディレクトリ(20-i)に格納される状態情報には、さらに割り込み可ライトロック状態情報とリクエストロック状態情報を含んでいる。

前記第1入出力コントローラが第I書き込み要求メッセージを発行する際に、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行する。

前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状态情報がフリー状態情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の強・弱書き込み要求メッセージ両方を受け付けることができないライトロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、第I書き込み許可メッセージを前記第1入出力コントローラに前記ネットワーク(2)を介して出力する。

前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状态情報がフリー状態情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができない割り込み可ライトロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、第I書き込み許可メッセージを前記第1入出力コントローラに前記ネットワーク(2)を介して出力する。

前記割り込み可ライトロック状態情報は、この状態にした弱書き込み要求メッセージを発行した入出力コントローラを特定する情報を含んでいる。

前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状态情報が前記ライトロック状態情報あるいはリクエストロック状態



情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに前記ネットワーク(2)を介して出力する。

前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状態情報が前記ライトロック状態に格納されている前記第Iデータの状态情報が前記割り込み可ライトロック状態情報、前記ライトロック状態情報、あるいはリクエストロック状態情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに前記ネットワーク(2)を介して出力する。

前記第1入出力コントローラは、前記第I不許可メッセージを受け取ったとき、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行する。

前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状态情報が前記割り込み可ライトロック状態情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)は、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記割り込み可ライトロック状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができないリクエストロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、第I再試行要求メッセージを前記ディレクトリ(20-i)に格納された入出力コントローラを特定する情報が指す第2入出力コントローラに宛てて前記ネットワーク(2)を介して出力する。

前記第I再試行要求メッセージを受け取った前記第2入出力コントローラは、再試行処理を行う。

#### 【0028】

本発明のマルチプロセッサシステムにおいて、前記第2入出力コントローラは以下のよう

に再試行処理を行う。  
第I開放メッセージを前記第1プロセッサノード(20-i)の前記メモリコントローラ(40-i)に前記ネットワーク(2)を介して出力する。

前記第Iデータとアドレスを同じくするライトメッセージで書き込み許可メッセージを受け取り済みのものに関して、まだ更新メッセージを発行していなければ更新メッセージの発行を停止して、前記第I開放メッセージの発行後に、前記第1プロセッサノード(20-i)に宛てて書き込み要求メッセージを発行する。

前記第I開放メッセージを受けた前記第1プロセッサノード(20-i)の前記メモリコントローラ(40-i)は、前記リクエストロック状態情報に代えてライトロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、第I書き込み許可メッセージを前記第1入出力コントローラに宛てて前記ネットワーク(2)を介して出力する。

#### 【0029】

本発明のメモリアクセス処理方法は、マルチプロセッサシステムに適用される。マルチプロセッサシステムは、複数のプロセッサノード(1-1~1-m)(mは1以上の整数である)と、複数の入出力ノード(3-1~3-n)(nは1以上の整数である)と、複数の入出力コントローラとを具備している。

前記複数のプロセッサノードの各々(1-i)(i=1, 2, ..., m)には、複数のプロセッサ(10-i-1, 10-i-2)と、複数のデータを格納する主記憶部(30-i)と、前記複数のデータの各々に対してアクセス要求を受け付けることが可能なフリー状態情報が格納されたディレクトリ(20-i)と、前記複数のプロセッサ(10-i-

1、10-i-2)と前記主記憶部(30-i)と前記ディレクトリ(20-i)とに接続されたメモリコントローラ(40-i)とが設けられている。前記複数の入出力ノードの各々(3-j)(j=1、2、…、n)には、ライトメッセージを発行する複数の入出力デバイス(60-j-1、60-j-2)が設けられている。

本発明のメモリアクセス処理方法は、前記複数の入出力ノード(3-1～3-n)のうちの第1入出力ノード(3-1)の前記複数の入出力デバイス(60-1-1、60-1-2)によって1番目からM番目(Mは1以上の整数である)までのM個のデータに対するM個のライトメッセージが発行されたとき、前記複数の入出力コントローラのうちの第1入出力コントローラが、前記M個のそれぞれのデータに対するM個のライトランザクションを開始し、前記M個のデータのうちの第Iデータ(Iは、I=1、2、…、Mを満たす整数の何れか)は、前記複数のプロセッサノード(1-1～1-m)のうちの第1プロセッサノード(1-i)をホームとするデータであり、第Iライトメッセージは前記第1プロセッサノード(1-i)の前記主記憶部(30-i)に格納された前記複数のデータのうちの第Iデータの値を第Iライトメッセージで指定される値に更新するための命令であり、前記第1入出力コントローラが、前記第Iライトランザクションの処理として、第I書き込み要求メッセージを前記第1プロセッサノード(1-i)に出力するステップと、

前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)が、前記第I書き込み要求メッセージを受け取ったとき、前記第Iデータに対して、前記フリー状態情報に代えて、前記第Iデータに対するプロセッサや入出力デバイスからの読み出し要求や他の書き込み要求メッセージを受け付けることができないライトロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、前記第I書き込み要求メッセージに対して第I書き込み許可メッセージを前記第1入出力コントローラに出力するステップと、前記第1入出力コントローラが、前記第I書き込み許可メッセージを受け取ったとき、第Iライトランザクションの処理として更新メッセージ発行処理を行うステップとを含む。

前記第1入出力コントローラが行う前記更新メッセージ発行処理は、第1から第I書き込み許可メッセージまでのI個の書き込み許可メッセージを既に受け取っているか否かを検査するステップと、前記I個の書き込み許可メッセージをまだ受け取っていないければ、前記第I更新メッセージ発行処理を終了するステップと、前記I個の書き込み許可メッセージを既に受け取っていれば、前記第Iライトメッセージで指定される値を含む第I更新メッセージを前記第1プロセッサノードに出力して第Iライトランザクションを完了させ、(I+1)がM以下であれば第(I+1)ライトランザクションの更新メッセージ発行処理を行い、(I+1)がMより大きければ前記第I更新メッセージ発行処理を終了するステップとを含む。

前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)が、前記第I更新メッセージを受け取ったとき、前記第Iデータに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納すると共に、前記第1プロセッサノード(1-i)の前記主記憶部(30-i)に格納された前記第Iデータの値を前記第I更新メッセージで指定される値に更新するステップとを含んでいる。

#### 【0030】

本発明のメモリアクセス処理方法は、更に、前記複数の入出力ノードの各々(3-j)の前記複数の入出力デバイス(60-j-1、60-j-2)によって発行される前記M個のライトメッセージを調停するステップを含んでいる。

#### 【0031】

本発明のメモリアクセス処理方法は、更に、前記第Iデータに対して前記ライトロック状態情報が前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されているときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)が、前記第Iデータに対する前記第I書き込み要求メッセージを前記第1入出力

コントローラから受け取った場合、前記第 I 書き込み要求メッセージに対して第 I 開放要求メッセージを前記第 1 入出力コントローラに出力するステップと、前記第 1 入出力コントローラが、前記第 I 開放要求メッセージを受けて前記第 I 書き込み要求メッセージを前記第 1 プロセッサノード (1-i) の前記メモリコントローラ (40-i) に出力すると共に、開放処理を行なうステップとを含んでいる。

前記開放処理を行なうステップは、前記第 I データに後続する第 K ライトトランザクション {K は、 $K = I + 1, I + 2, \dots, M$  を満たす整数であり、 $I + 1$  は、 $I < (I + 1) < M$  を満たす整数であり、 $I + 2$  は、 $(I + 1) < (I + 2) < M$  を満たす整数である} の進捗を検査し、未だ第 K 書き込み要求メッセージを発行していない場合、第 K 書き込み要求メッセージの発行を停止するステップと、既に第 K 書き込み要求メッセージを発行し第 K 書き込み許可メッセージを受け取っている場合、第 K 開放メッセージを前記第 K データのホームである第 2 プロセッサノード (1-k) の前記メモリコントローラ (40-k) に出力するステップと、既に第 K 書き込み要求メッセージを発行しまだ第 K 書き込み許可メッセージを受け取っていない場合は、第 K 書き込み許可メッセージを受け取った時点で前記第 K 開放メッセージの発行を行なうステップとを含む。

前記メモリアクセス処理方法は、更に、前記第 2 プロセッサノード (1-k) の前記メモリコントローラ (40-k) が、前記第 K 開放メッセージを受け取ったとき、前記第 I データに対して前記ライトロック状態情報に代えて前記フリー状態情報を前記第 2 プロセッサノード (1-k) の前記ディレクトリ (20-k) に格納するステップを含んでいる。

#### 【0032】

本発明のメモリアクセス処理方法において、前記ディレクトリ (20-i) に格納される状態情報には、さらに割り込み可ライトロック状態情報とリクエストロック状態情報を含んでいる。

本発明のメモリアクセス処理方法は、更に、前記第 1 入出力コントローラが第 I 書き込み要求メッセージを発行する際に、第 1 から第 (I-1) 書き込み許可メッセージまでの (I-1) 個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第 I 強書き込み要求メッセージを、受け取り済みでなければ第 I 弱書き込み要求メッセージを発行するステップと、前記第 1 プロセッサノード (1-i) の前記ディレクトリ (20-i) に格納されている前記第 I データの状態情報がフリー状態情報であるときに、前記第 1 プロセッサノード (1-i) の前記メモリコントローラ (40-i) が、前記第 I 強書き込み要求メッセージを前記第 1 入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第 I データに対する他の強・弱書き込み要求メッセージ両方を受け付けることができないライトロック状態情報を前記第 1 プロセッサノード (1-i) の前記ディレクトリ (20-i) に格納し、第 I 書き込み許可メッセージを前記第 1 入出力コントローラに出力するステップと、前記第 1 プロセッサノード (1-i) の前記ディレクトリ (20-i) に格納されている前記第 I データの状態情報がフリー状態情報であるときに、前記第 1 プロセッサノード (1-i) の前記メモリコントローラ (40-i) が、前記第 I 弱書き込み要求メッセージを前記第 1 入出力コントローラから受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第 I データに対する他の弱書き込み要求メッセージを受け付けることができない割り込み可ライトロック状態情報を前記第 1 プロセッサノード (1-i) の前記ディレクトリ (20-i) に格納し、第 I 書き込み許可メッセージを前記第 1 入出力コントローラに出力するステップと、前記割り込み可ライトロック状態情報は、この状態にした弱書き込み要求メッセージを発行した入出力コントローラを特定する情報を含んでいる。

本発明のメモリアクセス処理方法は、更に、前記第 1 プロセッサノード (1-i) の前記ディレクトリ (20-i) に格納されている前記第 I データの状態情報が前記ライトロック状態情報あるいはリクエストロック状態情報であるときに、前記第 1 プロセッサノード (1-i) の前記メモリコントローラ (40-i) が、前記第 I 強書き込み要求メッセ

ージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに出力するステップと、前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状態情報が前記ライトロック状態に格納されている前記第Iデータの状態情報が前記割り込み可ライトロック状態情報、前記ライトロック状態情報、あるいはリクエストロック状態情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)が、前記第I弱書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、第I不許可メッセージを前記第1入出力コントローラに出力するステップと、前記第1入出力コントローラが、前記第I不許可メッセージを受け取ったとき、第1から第(I-1)書き込み許可メッセージまでの(I-1)個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行するステップと、前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納されている前記第Iデータの状態情報が前記割り込み可ライトロック状態情報であるときに、前記第1プロセッサノード(1-i)の前記メモリコントローラ(40-i)が、前記第I強書き込み要求メッセージを前記第1入出力コントローラから受け取った場合、前記割り込み可ライトロック状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができないリクエストロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、第I再試行要求メッセージを前記ディレクトリ(20-i)に格納された入出力コントローラを特定する情報が指す第2入出力コントローラに宛てて出力するステップと、前記第I再試行要求メッセージを受け取った前記第2入出力コントローラが、再試行処理を行うステップとを含んでいる。

#### 【0033】

本発明のメモリアクセス処理方法において、前記第2入出力コントローラが再試行処理を行うステップは、第I開放メッセージを前記第1プロセッサノード(20-i)の前記メモリコントローラ(40-i)に出力するステップと、前記第Iデータとアドレスを同じくするライトメッセージで書き込み許可メッセージを受け取り済みのものに関して、まだ更新メッセージを発行していなければ更新メッセージの発行を停止して、前記第I開放メッセージの発行後に、前記第1プロセッサノード(20-i)に宛てて書き込み要求メッセージを発行するステップと、前記第I開放メッセージを受けた前記第1プロセッサノード(20-i)の前記メモリコントローラ(40-i)は、前記リクエストロック状態情報に代えてライトロック状態情報を前記第1プロセッサノード(1-i)の前記ディレクトリ(20-i)に格納し、第I書き込み許可メッセージを前記第1入出力コントローラに宛てて出力するステップとを含んでいる。

#### 【発明の効果】

#### 【0034】

以上の説明により、本発明のマルチプロセッサシステム及びメモリアクセス処理方法によれば、入出力コントローラが、異なるプロセッサノードを宛先とする入出力デバイスからの複数のライトメッセージを連続して処理することができる。

本発明のマルチプロセッサシステム及びメモリアクセス処理方法によれば、入出力コントローラが複数のライトメッセージを連続して処理するため、入出力コントローラがライトメッセージの処理を行う場合に要する時間を従来のそれよりも短くすることができる。

#### 【発明を実施するための最良の形態】

#### 【0035】

以下に添付図面を参照して、本発明のマルチプロセッサシステムについて詳細に説明する。

#### 【0036】

図4にマルチプロセッサシステムの構成を示す。本発明のマルチプロセッサシステムは、複数のプロセッサノード1-1~1-m(mは1以上の整数)と、複数の入出力ノード

3-1~3-n (nは1以上の整数)とを具備している。複数のプロセッサノード1-1~1-mと、複数の入出力ノード3-1~3-nとは、ネットワーク2に接続され、外部からのクロックに応じて動作する。プロセッサノード1-i (i=1, 2, ..., m) には、複数のプロセッサ10-i-1, 10-i-2と、ディレクトリ20-iと、主記憶部(メモリ)30-iと、メモリコントローラ40-iとが設けられている。メモリコントローラ40-iは、プロセッサ10-i-1, 10-i-2とディレクトリ20-iと主記憶部30-iとに接続されている。入出力ノード3-j (j=1, 2, ..., n) には、入出力コントローラ50-jと、外部からの命令によりメッセージを発行する複数の入出力デバイス60-j-1, 60-j-2とが設けられている。メッセージは、コマンドの種類を表すコマンド種別と、アドレスとを含み、例えば、メッセージがライトメッセージである場合、コマンド種別はライトを表す。また、ネットワーク2はメッセージの配送を行い、2点間のメッセージの順序を保障する。

#### 【0037】

主記憶部30-iには、複数のデータが格納されている。複数のデータの各々は、その内容を表す値を含んでいる。

ディレクトリ20-iは、主記憶部30-iに格納されている各データの一貫性制御に関する情報を、例えば128バイトのブロック単位、で保持している。各ブロックの情報としては、一貫性制御に関する状態情報を含む。この状態情報には、アクセス要求を受け付けることができるフリー状態情報と、他のアクセス要求を受け付けることができないライトロック状態情報が含まれる。

ここで、一貫性制御(一貫性処理)について説明する。マルチプロセッサシステムでは、複数のプロセッサが存在する。また、複数のプロセッサがそれぞれキャッシュを具備し、データのコピーを保持する。そのため、ひとつのデータについて、メモリの値と、コピーを取った複数のキャッシュ値とを一致させる処理が必要になる。この一致させる処理としては、コピーを無効化する処理が例示される。このように、データの値を一致させる、即ち、データの一貫性を取ることを一貫性制御(一貫性処理)と呼ぶ。

データの一貫性制御に関する情報については、後述の実施例にて説明する。

#### 【0038】

本発明は、入出力コントローラ50-jとメモリコントローラ40-iの間での一連のメッセージを工夫することで、ホームを異にするライトメッセージの連続発行を可能とするものである。

いま、入出力ノード3-1~3-nのうちの第1入出力ノード(入出力ノード3-1とする)の入出力デバイス60-1-1, 60-1-2)によって1番目からM番目(Mは1以上の整数)までのM個のデータに対するM個のライトメッセージが発行されたものとする。このとき、複数の入出力コントローラのうちの第1入出力コントローラ(入出力ノード3-1の入出力コントローラ50-1)がこれらM個のライトメッセージを受け取ると、前記M個のそれぞれのデータに対するM個のライトトランザクションを開始する。

M個のデータのうちの第Iデータ(I=1, 2, ..., M)は、プロセッサノード1-1~1-mのうちの1つのプロセッサノード(プロセッサノード1-iとする)をホームとするデータである。M個のライトメッセージのうちの第Iライトメッセージは、プロセッサノード1-iの主記憶部30-iに格納された複数のデータのうちの第Iデータの値を第Iライトメッセージで指定される値に更新するための命令である。以降、第Iライトメッセージを例にとり説明する。

#### 【0039】

第Iライトメッセージを受けた入出力コントローラ50-1は、第Iライトトランザクションを開始して、第I書き込み要求メッセージを、ネットワーク2を介してプロセッサノード1-iに出力する。プロセッサノード1-iのメモリコントローラ40-iは、第I書き込み要求メッセージを受け取ったとき、第Iデータに対して、フリー状態情報に代えて、第Iデータに対する他の書き込み要求メッセージを受け付けることができないライトロック状態情報をプロセッサノード1-iのディレクトリ20-iに格納し、第I書き

込み要求メッセージに対して第I書き込み許可メッセージを入出力ノード3-1にネットワーク2を介して出力する。

【0040】

入出力ノード3-1の入出力コントローラ50-1は、第I書き込み許可メッセージを受け取ったとき、図14に示されるような発行処理を実行する。

【0041】

まず、入出力コントローラ50-1は、第1から第I書き込み許可メッセージまでのI個の書き込み許可メッセージを既に受け取っているか否か进行检查する(図14のステップS1)。ここで、入出力コントローラ50-1は、上記のI個の書き込み許可メッセージを未だ受け取っていない場合(図14のステップS1-NO)、発行処理を終了し、次の書き込み許可メッセージの到着を待つ。入出力コントローラ50-1は、上記のI個の書き込み許可メッセージを既に受け取っている場合(図14のステップS1-YES)、第Iライトメッセージで指定される値を含む第I更新メッセージをプロセッサノード1-iにネットワーク2を介して出力する(図14のステップS2)。

【0042】

次に、入出力コントローラ50-1は、 $I = I + 1$ とし、IがM以下であるかどうか进行检查する(図14のステップS3、S4)。IがM以下である、即ち、第Iライトメッセージを受け取って第Iライトトランザクションを開始していれば(図14のステップS4-YES)、ステップS1を実行する。IがMより大きい、即ち、まだ第Iライトメッセージを受け取っていないければ(図14のステップS4-NO)処理を終了する。

【0043】

このように、入出力コントローラ50-1は、ステップS1~S4を繰り返し実行し、当該ライトトランザクションおよび先行するライトトランザクションが全て書き込み許可メッセージを受け取っていれば、更新メッセージを発行する。

【0044】

プロセッサノード1-iのメモリコントローラ40-iは、第I更新メッセージを受け取ったとき、第Iデータに対してライトロック状態情報に代えてフリー状態情報をプロセッサノード1-iのディレクトリ20-iに格納すると共に、プロセッサノード1-iの主記憶部30-iに格納された第Iデータの値を第I更新メッセージで指定される値に更新する。

【0045】

このように、本発明のマルチプロセッサシステムによれば、入出力コントローラ50-jが、異なるプロセッサノード1-iを宛先とする入出力デバイス60-j-1、60-j-2からの複数のライトメッセージを連続して処理することができる。

【0046】

図6を参照しながら、順序制約のあるライトメッセージが連続して発行された場合の動作を具体的に説明する。図は、入出力デバイス60-1-1が、M個(M=3)のデータとしてデータA、B、Cに対するライトメッセージをそれぞれステップ1、2、3で発行した場合の動作を示している。ここで、データAおよびデータBはプロセッサノード1-1をホームとし、データCはプロセッサノード1-2をホームとするデータであるとする。また、1ステップは1クロックに対応する。

【0047】

入出力コントローラ50-1は、ステップ2にて、第1ライトメッセージとしてライトAメッセージを受け取る。このとき、入出力コントローラ50-1は、ライトAトランザクションを開始し、ステップ3にて、ホームのメモリコントローラ40-1に宛ててネットワーク2に、第1書き込み要求メッセージとして書き込み要求Aメッセージを出力する。

メモリコントローラ40-1は、ステップ4にて、入出力コントローラ50-1からの書き込み要求Aメッセージを受け取ったとき、ディレクトリ20-1が保持するデータAの状態情報をフリー状態情報からライトロック状態情報に更新する。また、メモリコント

ローラ40-1は、ステップ5にて、入出力コントローラ50-1に宛ててネットワーク2に、第1書き込み許可メッセージとして書き込み許可Aメッセージを出力する。

入出力コントローラ50-1は、ステップ6にて、書き込み許可Aメッセージを受け取る。このとき、先行するライトメッセージが存在しないので、入出力コントローラ50-1は、ステップ7にて、第1ライトメッセージで指定される値を含む第1更新メッセージとして、更新Aメッセージをメモリコントローラ40-1に宛ててネットワーク2に出力する。

メモリコントローラ40-1は、ステップ8にて、入出力コントローラ50-1からの更新Aメッセージを受け取る。このとき、メモリコントローラ40-1は、ディレクトリ20-1が保持するデータAの状態情報をライトロック状態情報からフリー状態情報に更新し、主記憶部30-1のデータの値を更新Aメッセージで指定される値に更新する（入出力コントローラ50-1からのライトAメッセージであるデータAを格納する）。

#### 【0048】

入出力コントローラ50-1は、ステップ3にて、第2ライトメッセージとしてライトBメッセージを受け取る。このとき、入出力コントローラ50-1は、ライトBトランザクションを開始し、ステップ4にて、ホムのメモリコントローラ40-1に宛ててネットワーク2に、第2書き込み要求メッセージとして書き込み要求Bメッセージを出力する。

メモリコントローラ40-1は、ステップ5にて、入出力コントローラ50-1からの書き込み要求Bメッセージを受け取ったとき、ディレクトリ20-1が保持するデータBの状態情報をフリー状態情報からライトロック状態情報に更新する。また、メモリコントローラ40-1は、ステップ6にて、入出力コントローラ50-1に宛ててネットワーク2に、第2書き込み許可メッセージとして書き込み許可Bメッセージを出力する。

入出力コントローラ50-1は、ステップ7にて、書き込み許可Bメッセージを受け取る。このとき、入出力コントローラ50-1は、先行するライトAトランザクションの進捗を検査する。入出力コントローラ50-1は、ステップ7より前のステップ6で既に第1書き込み許可メッセージである書き込み許可Aメッセージを受け取っている。よって、入出力コントローラ50-1は、ステップ8にて、第2ライトメッセージで指定される値を含む第2更新メッセージとして、更新Bメッセージをメモリコントローラ40-1に宛ててネットワーク2に出力する。

メモリコントローラ40-1は、ステップ9にて、入出力コントローラ50-1からの更新Bメッセージを受け取る。このとき、メモリコントローラ40-1は、ディレクトリ20-1が保持するデータBの状態情報をライトロック状態情報からフリー状態情報に更新し、主記憶部30-1のデータの値を更新Bメッセージで指定される値に更新する（入出力コントローラ50-1からのライトBメッセージであるデータBを格納する）。

#### 【0049】

入出力コントローラ50-1は、ステップ4にて、第3ライトメッセージとしてライトCメッセージを受け取る。このとき、入出力コントローラ50-1は、ライトCトランザクションを開始し、ステップ5にて、ホムのメモリコントローラ40-2に宛ててネットワーク2に、第3書き込み要求メッセージとして書き込み要求Cメッセージを出力する。

メモリコントローラ40-2は、ステップ6にて、入出力コントローラ50-1からの書き込み要求Cメッセージを受け取ったとき、ディレクトリ20-2が保持するデータCの状態情報をフリー状態情報からライトロック状態情報に更新する。また、メモリコントローラ40-1は、ステップ7にて、入出力コントローラ50-1に宛ててネットワーク2に、第3書き込み許可メッセージとして書き込み許可Cメッセージを出力する。

入出力コントローラ50-1は、ステップ8にて、書き込み許可Cメッセージを受け取る。このとき、入出力コントローラ50-1は、先行するライトAトランザクションおよびライトBトランザクションの進捗を検査する。入出力コントローラ50-1は、ステップ8より前のステップ6で既に書き込み許可Aメッセージを受け取り、ステップ8より前



のステップ7で既書き込み許可Bメッセージを受け取っている。よって、入出力コントローラ50-1は、ステップ9にて、第3ライトメッセージで指定される値を含む第3更新メッセージとして、更新Cメッセージをメモリコントローラ40-2に宛ててネットワーク2に出力する。

メモリコントローラ40-2は、ステップ10にて、入出力コントローラ50-1からの更新Cメッセージを受け取る。このとき、メモリコントローラ40-2は、ディレクトリ20-2が保持するデータCの状態情報をライトロック状態情報からフリー状態情報に更新し、主記憶部30-2のデータの値を更新Cメッセージで指定される値に更新する（入出力コントローラ50-1からのライトCメッセージであるデータCを格納する）。

#### 【0050】

入出力コントローラ50-jには、図5に示されるように、セクタ71-jと、メッセージ格納キュー72-jと、ライトポイント73-jと、リードポイントA74-jと、リードポイントB75-jとが設けられている。入出力コントローラ50-jは、複数の入出力デバイス60-j-1、60-j-2からのメッセージをセクタ71-jで調停してメッセージ格納キュー72-jに書き込む。メッセージ格納キュー72-jにおける書き込みの制御はライトポイント73-jを用いて行われ、読み出しの制御はリードポイントA74-jとリードポイントB75-jの二つを用いて行われる。また、入出力コントローラ50-jは、ネットワーク2から送られてくるメッセージに応じて許可フラグ76-1をメッセージ格納キュー72-jに設定（格納）し、メッセージ格納キュー72-1の制御に該許可フラグ76-1も用いられる。

ここで、上記図6に示した動作フローでの入出力コントローラ50-1のメッセージ格納キュー72-1、ライトポイント73-1、リードポイントA74-1、リードポイントB75-1、許可フラグ76-1の値の遷移を図7A～図7Iに示す。

#### 【0051】

図7Aに示されるように、初期状態として、ステップ1にて、ライトポイント73-1、リードポイントA74-1、リードポイントB75-1は全て「0」を指している。

#### 【0052】

入出力コントローラ50-1は、ステップ2にて、入出力デバイス60-1-1からライトAメッセージを受け取ると、図7Bに示されるように、メッセージ格納キュー72-1のライトポイント73-1が指すエントリ「0」にライトAメッセージを書き込み、同じく許可フラグ76-1のエントリ「0」の値を「0」に設定し、ライトポイント73-1を「1」に更新する。

#### 【0053】

入出力コントローラ50-1は、ステップ3にて、入出力デバイス60-1-1からライトBメッセージを受け取ると、図7Cに示されるように、メッセージ格納キュー72-1のライトポイント73-1が指すエントリ「1」にライトBメッセージを書き込み、同じく許可フラグ76-1のエントリ「1」の値を「0」に設定し、ライトポイント73-1を「2」に更新する。

また、ステップ2にてメッセージ格納キュー72-1に有効なエントリが存在することになる。このため、図7Cに示されるように、入出力コントローラ50-1は、ステップ3にて、リードポイントA74-1が指すエントリ「0」の情報により、書き込み要求Aメッセージをネットワーク2に出力し、リードポイントA74-1の値を「1」に更新する。

#### 【0054】

入出力コントローラ50-1は、ステップ4にて、入出力デバイス60-1-1からライトCメッセージを受け取る。このとき、図7Dに示されるように、入出力コントローラ50-1は、メッセージ格納キュー72-1のライトポイント73-1が指すエントリ「2」にライトCメッセージを書き込み、同じく許可フラグ76-1のエントリ「2」の値を「0」に設定し、ライトポイント73-1を「3」に更新する。

また、ステップ3にてメッセージ格納キュー72-1に有効なエントリが存在すること



になる。このため、図7Dに示されるように、入出力コントローラ50-1は、ステップ4にて、リードポインタA74-1が示すエントリ「1」の情報により、書き込み要求Bメッセージをネットワーク2に出力し、リードポインタA74-1の値を「1」から「2」に更新する。

**【0055】**

ステップ4にてメッセージ格納キュー72-1に有効なエントリが存在することになる。このため、図7Eに示されるように、入出力コントローラ50-1は、ステップ5にて、リードポインタA74-1が示すエントリ「2」の情報により、書き込み要求Cメッセージをネットワーク2に出力し、リードポインタA74-1の値を「2」から「3」に更新する。

**【0056】**

入出力コントローラ50-1は、ステップ6にて、書き込み許可Aメッセージを受け取ったとき、図7Fに示されるように、許可フラグ76-1のデータAに該当するエントリ「0」の値を「0」から「1」に更新する。そしてリードポインタB75-1が指すエントリと書き込み許可メッセージを受け取ったデータAが格納されているエントリが一致しているかどうかを検査する。この値が一致するということは、先行するライトトランザクションが存在しない、あるいは既に書き込み許可メッセージを受け取って更新メッセージを発行し完了していることを示している。ここでは「0」で一致するので、更新メッセージの発行処理を行う。次の書き込み許可メッセージの到着を待つ。

**【0057】**

入出力コントローラ50-1は、ステップ6にてリードポインタB75-1が指すメッセージ格納キュー72-1のエントリ「0」の値を読み出し、ステップ7にて、ネットワーク2に更新Aメッセージを出力する。このとき、図7Gに示されるように、入出力コントローラ50-1は、リードポインタB75-1の値を「0」から「1」に更新する。

次に、リードポインタB75-1とライトポインタ73-1の値を比較し、リードポインタB75-1の値が小さく未完了の書き込みが存在することを示す場合、未完了のライトトランザクションで既に書き込み許可メッセージを受け取り済みのものが存在するかどうかを検査する。ここでは、リードポインタB75-1の値「1」はライトポインタ73-1の値「3」より小さいので、リードポインタB75-1が指す許可フラグ76-1のエントリ「1」の値を読み出す。値は「0」であり書き込み許可メッセージをまだ受け取っていないことを示すので、エントリ「1」の更新メッセージ発行処理は行わず次の書き込み許可メッセージの到着を待つ。

**【0058】**

入出力コントローラ50-1は、ステップ7にて、許可Bメッセージを受け取ると、図7Gに示されるように、許可フラグ76-1のデータBに該当するエントリ「1」の値を「0」から「1」に更新する。そしてリードポインタB75-1が指すエントリと書き込み許可メッセージを受け取ったデータBが格納されているエントリが一致しているかどうかを検査する。ここでは「1」で一致するので、更新メッセージの発行処理を行う。

**【0059】**

入出力コントローラ50-1は、ステップ7にてリードポインタB75-1が指すメッセージ格納キュー72-1のエントリ「1」の値を読み出し、ステップ8にて、ネットワーク2にライトBメッセージを出力する。このとき、図7Hに示されるように、入出力コントローラ50-1は、リードポインタB75-1の値を「1」から「2」に更新する。

次に、リードポインタB75-1の値「2」とライトポインタ73-1の値「3」を比較すると、リードポインタB75-1の値が小さいので、リードポインタB75-1が指す許可フラグ76-1のエントリ「2」の値を読み出す。値は「0」であり書き込み許可メッセージをまだ受け取っていないことを示すので、エントリ「2」の更新メッセージ発行処理は行わず次の書き込み許可メッセージの到着を待つ。

**【0060】**

入出力コントローラ50-1は、ステップ8にて、書き込み許可Cメッセージを受け取

ると、図7Hに示されるように、許可フラグ76-1のデータCに該当するエントリ「2」の値を「0」から「1」に更新する。そしてリードポインタB75-1が指すエントリと書き込み許可メッセージを受け取ったデータCが格納されているエントリが一致しているかどうかを検査する。ここでは「2」で一致するので、更新メッセージの発行処理を行う。

#### 【0061】

入出力コントローラ50-1は、ステップ8にてリードポインタB75-1が指すメッセージ格納キュー72-1のエントリ「2」の値を読み出し、ステップ9にて、ネットワーク2に更新Cメッセージを出力する。このとき、図7Iに示されるように、入出力コントローラ50-1は、リードポインタB75-1の値を「2」から「3」に更新する。

次に、リードポインタB75-1の値「3」とライトポインタ73-1の値「3」を比較すると、値が一致し未完了のライトトランザクションが存在しないことを示すので、処理を終了する。

#### 【0062】

上記のように動作することで、ホームを同じにするライトトランザクションの順序も、異なるホームへのライトトランザクションの順序も保障することができる。

ライトBトランザクションとライトAトランザクションの順序関係は次のような理由で保障される。更新Bメッセージを発行する時点で、既に書き込み許可Aメッセージを受け取り済みであり、ホームのディレクトリ20-1の状態情報が、ライトロック状態に遷移していることが保障される。ライトロック状態に遷移していれば、プロセッサや入出力デバイスは更新Aメッセージを受け取ってデータAの値が更新されフリー状態に遷移した後の更新された値しか読み出すことができない。よって、データBの更新された値が読み出せる時点で、データAの更新された値しか読み出すことができないので、順序が保障されたことになる。

#### 【0063】

ライトCトランザクションとライトBトランザクションの順序関係も同じな理由で保障される。更新Cメッセージを発行する時点で、既に書き込み許可Bメッセージを受け取り済みであり、ホームのディレクトリ20-1の状態情報が、ライトロック状態に遷移していることが保障される。ライトロック状態に遷移していれば、プロセッサや入出力デバイスは更新Bメッセージを受け取ってデータBの値が更新されフリー状態に遷移した後の更新された値しか読み出すことができない。よって、データCの更新された値が読み出せる時点で、データBの更新された値しか読み出すことができないので、順序が保障されたことになる。

#### 【0064】

また、上記処理は10ステップで完了しており、従来技術と比べて入出力デバイス60-j-1からのライトメッセージの処理性能を向上させることができる。

#### 【0065】

図8に示すように、上記の入出力コントローラ50-jは、入出力ノード3-j内に設けられているが、入出力コントローラ50-jとして入出力コントローラ52-jがネットワーク2に接続されていてもよい。この場合、入出力ノード3-jには、上記の入出力セクタ71-jに対応する入出力セクタ51-jが設けられていることが好ましい。入出力セクタ51-jは、複数の入出力デバイス60-j-1、60-j-2が出力するメッセージを調停してネットワーク2に出力し、入出力コントローラ52-jは、ネットワーク2に出力されたメッセージを受けて、図4に示した入出力コントローラ50-jと同様の処理を行う構成でも良い。

#### 【0066】

また、上記図8に示した入出力コントローラ52-jは、ネットワーク2に接続されるのではなく、入出力コントローラ52-i（図示しない）としてプロセッサノード1-i内に設けられていても良い。この場合、入出力ノード3-jには、上記の入出力セクタ51-jが設けられていることが好ましい。

## 【0067】

また、図4や図8に示す構成で、 $m$ と $n$ とが等しく、プロセッサノード $1-i$ と入出力ノード $3-j$ とで一つのノード（図示しない）を構成してもよい。この場合、そのノードには、ネットワーク2が接続され、上記の入出力セクタ $51-j$ が設けられていることが好ましい。

## 【0068】

以上の説明により、本発明のマルチプロセッサシステムによれば、入出力コントローラ $50-j$ が、異なるプロセッサノード $1-i$ を宛先とする入出力デバイス $60-j-1$ 、 $60-j-2$ からの複数のライトメッセージを連続して処理することができる。

本発明のマルチプロセッサシステムによれば、入出力コントローラ $50-j$ が複数のライトメッセージを連続して処理するため、入出力コントローラ $50-j$ がライトメッセージの処理を行う場合に要する時間を従来のそれよりも短くすることができる。

## 【0069】

以降、入出力コントローラ $50-j$ とメモリコントローラ $40-i$ 間のやりとりをより詳細に説明する。

## 【実施例1】

## 【0070】

上述のように、ディレクトリ $20-i$ は、主記憶部 $30-i$ に格納されているデータの一貫性制御に関する情報を、例えば128バイトのブロック単位で保持している。データの一貫性制御に関する情報は、主記憶部 $30-i$ の各ブロックの状態情報（データの一貫性制御に関する状態情報）と、マップ情報とを含んでいる。

## 【0071】

格納されるブロックの状態情報（データの一貫性制御に関する状態情報）には、上述のように、他のアクセス要求を受け付けることができるフリー状態情報と、他のアクセス要求を受け付けることができないライトロック状態情報が含まれる。

フリー状態情報は、例えば、Uncached、Clean、Dirty（以降U、C、Dとも略す）の3つの状態情報からなる。

Uncachedは、複数のプロセッサノード $1-1 \sim 1-m$ のうちのどのプロセッサノードもデータをキャッシングしていないことを示す。

Cleanは、複数のプロセッサノード $1-1 \sim 1-m$ のうちの少なくとも1つのプロセッサノードがデータをキャッシングしていることを示す。

Dirtyは、複数のプロセッサノード $1-1 \sim 1-m$ のうちのある1つのプロセッサノードがデータをキャッシングし、最新のデータはその1つのプロセッサノードにのみ存在することを示す。

## 【0072】

また、データの一貫性制御に関する状態情報には、他のアクセス要求を受け付けることができない状態情報として、要求（リクエスト）されたメッセージのみを受け付けることができるリクエストロック状態情報が含まれるものとする。本発明において、ライトロック状態情報は、このリクエストロック状態情報と同一にしてもよいし異なる二つの状態情報としても良い。以降の説明では二つの異なる状態情報（以降リクエストロック状態情報をR、ライトロック状態情報をWとも略す）として存在する場合を例にとり説明する。

## 【0073】

また、マップ情報は、各プロセッサノード $1-i$ が該ブロックのデータをキャッシングしているかどうかを示す情報である。プロセッサノード1数分のビットを用いて、マップ情報について説明する。図4の構成例として $m$ を3とする。即ち、プロセッサノードが3ノード存在するので（プロセッサノード $1-1$ 、 $1-2$ 、 $1-3$ ）、3ビットで表現するものとする。

例えば“000”はどのプロセッサノード $1-1$ 、 $1-2$ 、 $1-3$ もデータをキャッシングしていないことを示す。“001”はプロセッサノード $1-1$ がデータをキャッシングしていることを示す。“010”はプロセッサノード $1-2$ がデータをキャッシングし

ていることを示す。“100”はプロセッサノード1-3がデータをキャッシングしていることを示す。同様に、“110”はプロセッサノード1-2とプロセッサノード1-3がデータをキャッシングしていることを示す。

#### 【0074】

図9を参照しながら、本発明の第1実施例に係るマルチプロセッサシステムの動作として、メモリコントローラ40-1が書き込み要求Aメッセージを受けたときに、ディレクトリ20-1の該当するブロックの状態情報がUncached状態情報であった場合の動作について説明する。図は、入出力デバイス60-1-1が、データAに対するライトメッセージをステップ1で発行した場合の動作を示している。ここで、データAはプロセッサノード1-1をホームとするデータであるとする。また、1ステップは1クロックに対応する。

#### 【0075】

入出力コントローラ50-1は、ステップ2にて、ライトAメッセージを受け取る。このとき、入出力コントローラ50-1は、ステップ3にて、ホームのメモリコントローラ40-1に宛ててネットワーク2に書き込み要求Aメッセージを出力する。

メモリコントローラ40-1は、ステップ4にて、入出力コントローラ50-1からの書き込み要求Aメッセージを受け取り、ディレクトリ20-1が保持するデータAの状態情報をフリー状態情報からライトロック状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“U、000”から“W、000”に更新する。ここで、“U、000”は状態情報がUであることを示し、マップ情報が“000”であることを示す。また、メモリコントローラ40-1は、ステップ5にて、入出力コントローラ50-1に宛ててネットワーク2に書き込み許可Aメッセージを出力する。

入出力コントローラ50-1は、ステップ6にて、書き込み許可Aメッセージを受け取る。次に、ステップ7にて、更新Aメッセージをメモリコントローラ40-1に宛ててネットワーク2に出力する。

メモリコントローラ40-1は、ステップ8にて、入出力コントローラ50-1からの更新Aメッセージを受け取り、主記憶部30-1のデータの値を更新Aメッセージで指定される値に更新する。そして、ディレクトリ20-1が保持するデータAの状態情報をライトロック状態情報からフリー状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“W、000”から“U、000”に更新する。

#### 【0076】

図10を参照しながら、本発明の第1実施例に係るマルチプロセッサシステムの動作として、メモリコントローラ40-1が書き込み要求Aメッセージを受けたときに、ディレクトリ20-1の該当するブロックの状態情報がClean状態情報で、プロセッサノード1-2とプロセッサノード1-3がデータをキャッシングしている場合の動作について説明する。図は、入出力デバイス60-1-1が、データAに対するライトメッセージをステップ1で発行した場合の動作を示している。ここで、データAはプロセッサノード1-1をホームとするデータであるとする。また、1ステップは1クロックに対応する。

#### 【0077】

入出力コントローラ50-1は、ステップ2にて、ライトAメッセージを受け取る。このとき、入出力コントローラ50-1は、ステップ3にて、ホームのメモリコントローラ40-1に宛ててネットワーク2に書き込み要求Aメッセージを出力する。

メモリコントローラ40-1は、ステップ4にて、入出力コントローラ50-1からの書き込み要求Aメッセージを受け取り、ディレクトリ20-1が保持するデータAの状態情報をフリー状態情報からライトロック状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“C、110”から“W、000”に更新する。ここで、“C、110”は状態情報がCであることを示し、マップ情報が“110”であることを示す。また、メモリコントローラ40-1は、ステップ5にて、入出力コントローラ50-1に宛ててネットワーク2に応答Aメッセージを出力し、メモリコントローラ40-2とメモリコントローラ40-3に宛てて無効化Aメッセージを出力する。このとき、応答Aメッセ

ージには、キャッシングしているプロセッサノードの数（この例では2）が付加される。

入出力コントローラ50-1は、ステップ6にて、応答Aメッセージを受け取る。また、メモリコントローラ40-2、メモリコントローラ40-3は、ステップ6にて、それぞれ無効化Aメッセージを受け取り、それぞれ当該プロセッサノード1-2、プロセッサノード1-3でキャッシングしているデータAを無効化する。そして、メモリコントローラ40-2、メモリコントローラ40-3は、ステップ7にて、それぞれネットワーク2に入出力コントローラ50-1に宛てて無効化完了Aメッセージを出力する。

入出力コントローラ50-1は、ステップ8にて、応答Aメッセージに付加されている数の無効化完了Aメッセージを受け取った時点で、書き込み許可Aメッセージを受け取ったと解釈（認識）する。そして、ステップ9にて、更新Aメッセージをメモリコントローラ40-1に宛ててネットワーク2に出力する。

メモリコントローラ40-1は、ステップ10にて、入出力コントローラ50-1からのライトAメッセージを受け取り、主記憶部30-1のデータの値を更新Aメッセージで指定される値に更新する。そして、ディレクトリ20-1が保持するデータAの状態情報をライトロック状態情報からフリー状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“W、000”から“U、000”に更新する。

#### 【0078】

図11を参照しながら、本発明の第1実施例に係るマルチプロセッサシステムの動作として、メモリコントローラ40-1が書き込み要求Aメッセージを受けたときに、ディレクトリ20-1の該当するブロックの状態情報がDirty状態情報で、プロセッサノード1-2がデータをキャッシングしている場合の動作について説明する。図は、入出力デバイス60-1-1が、データAに対するライトメッセージをステップ1で発行した場合の動作を示している。ここで、データAはプロセッサノード1-1をホームとするデータであるとする。また、1ステップは1クロックに対応する。

#### 【0079】

入出力コントローラ50-1は、ステップ2にて、ライトAメッセージを受け取る。このとき、入出力コントローラ50-1は、ステップ3にて、ホームのメモリコントローラ40-1に宛ててネットワーク2に書き込み要求Aメッセージを出力する。

メモリコントローラ40-1は、ステップ4にて、入出力コントローラ50-1からの書き込み要求Aメッセージを受け取り、ディレクトリ20-1が保持するデータAの状態情報をフリー状態情報からリクエストロック状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“D、010”から“R、010”に更新する。ここで、“R、010”は状態情報がRであることを示し、マップ情報が“010”であることを示す。また、メモリコントローラ40-1は、ステップ5にて、入出力コントローラ50-1に宛ててネットワーク2に書き戻し要求Aメッセージを出力する。ここで、リクエストロック状態情報である“R、010”への更新ではなく、ライトロック状態情報である“W、000”への更新であっても構わない。

メモリコントローラ40-2は、ステップ6にて、メモリコントローラ40-1からの書き戻し要求Aメッセージを受け取り、該プロセッサノード1-2でキャッシングしているデータAの書き戻しを行い、ステップ7にて、メモリコントローラ40-1に宛ててネットワーク2に書き戻しAメッセージを出力する。

メモリコントローラ40-1は、ステップ8にて、リクエストロック状態情報によりリクエストされたメッセージとして、書き戻しAメッセージをメモリコントローラ40-2から受け取り、ディレクトリ20-1が保持するデータAの状態情報をリクエストロック状態情報からライトロック状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“R、010”から“W、000”に更新する。また、メモリコントローラ40-1は、ステップ9にて、入出力コントローラ50-1に宛ててネットワーク2に書き込み許可Aメッセージを出力する。

入出力コントローラ50-1は、ステップ10にて、書き込み許可Aメッセージを受け取る。次に、ステップ11にて、更新Aメッセージをメモリコントローラ40-1に宛て

てネットワーク2に出力する。

メモリコントローラ40-1は、ステップ11にて、入出力コントローラ50-1からの更新Aメッセージを受け取り、主記憶部30-1のデータの値を更新Aメッセージで指定される値に更新する。そして、ディレクトリ20-1が保持するデータAの状態情報をライトロック状態情報からフリー状態情報に更新する。即ち、ディレクトリ20-1の該ブロックの値を“W、000”から“U、000”に更新する。

#### 【0080】

図12を参照しながら、本発明の第1実施例に係るマルチプロセッサシステムの動作として、メモリコントローラ40-1が書き込み要求Aメッセージを受けたときに、ディレクトリ20-1の該当するブロックの状態情報がリクエストロック状態情報あるいはライトロック状態情報であった場合の動作について説明する。図は、入出力デバイス60-1-1が、データAに対するライトメッセージをステップ1で発行した場合の動作を示している。ここで、データAはプロセッサノード1-1をホームとするデータであるとする。また、1ステップは1クロックに対応する。

#### 【0081】

入出力コントローラ50-1は、ステップ2にて、ライトAメッセージを受け取る。このとき、入出力コントローラ50-1は、ステップ3にて、ホームのメモリコントローラ40-1に宛ててネットワーク2に書き込み要求Aメッセージを出力する。

メモリコントローラ40-1は、ステップ4にて、入出力コントローラ50-1からの書き込み要求Aメッセージを受け取る。このとき、ディレクトリ20-1が保持するデータAの状態情報は、リクエストロック状態情報“R”あるいはライトロック状態情報“W”である。このため、メモリコントローラ40-1は、ステップ5にて、入出力コントローラ50-1に宛ててネットワーク2に不許可Aメッセージを出力する。

入出力コントローラ50-1は、ステップ6にて、不許可Aメッセージを受け取り、ステップ7にて、メモリコントローラ40-1に宛ててネットワーク2に書き込み要求Aメッセージを再出力する。ステップ7以降の動作は、これまで図9から図12を参照しながら説明した動作と同じであるので省略する。

#### 【0082】

以上のように動作することで、ディレクトリのマップ情報に入出力ノードの分のビットを加えなくて済む。また、ホームが受けたプロセッサからのアクセス要求を、入出力コントローラ50-jに対して転送せずに済み、メモリコントローラや入出力コントローラの構成が複雑にならずに済む。

#### 【実施例2】

#### 【0083】

本発明のマルチプロセッサシステムの入出力コントローラ50-jは図13に示すように構成されても良い。この入出力コントローラ50-jは、更に、開放処理ポインタ78-jと開放処理フラグ79-jを有する。初期状態で、開放処理フラグ79-jの値は“0”である。

基本動作は実施例1と同様であるので、ここでは異なる部分のみを説明し、同じ部分は省略する。また、実施の最良の形態と同様に、入出力ノード3-1～3-nのうちの第1入出力ノード（入出力ノード3-1とする）の入出力デバイス60-1-1、60-1-2によって1番目からM番目（Mは1以上の整数）までのM個のデータに対するM個のライトメッセージが発行されたものとする。以降、第Iライトメッセージを例にとり説明する。

#### 【0084】

メモリコントローラ40-iは、第Iデータに対して前記ライトロック状態情報が前記ディレクトリ20-iに格納されているときに、前記第Iデータに対する第I書き込み要求メッセージを前記入出力コントローラ50-1から受け取った場合、第I書き込み要求メッセージに対して第I開放要求メッセージを入出力コントローラ50-1にネットワーク2を介して出力する。第I開放要求メッセージを受けた入出力コントローラ50-1は

、第I書き込み要求メッセージをメモリコントローラ40-iに宛ててネットワーク2に出力する。また、開放処理フラグ79-1の値を“0”から“1”に更新し、開放処理ポインタ78-1の値を、第I番目のライトランザクションであることを示す情報“I”に設定する。そして後述の開放処理を行う。

【0085】

この開放処理フラグ79-1は、入出力コントローラが後述の開放処理を行っている状態であるかどうかを示し、開放処理ポインタは、その開放処理がどのライトランザクションによって引き起こされたものであるかを示す。開放処理ポインタ78-1の値は、開放処理を引き起こしたライトランザクションと他のライトランザクションとの順序関係（先行するものか後続するものか）を判断するのにも用いる。

【0086】

入出力コントローラ50-1は、開放処理フラグ79-1の値が“1”の間、第Iライトに後続する第Kランザクション  $\{Kは、K=I+1、I+2、…、Mを満たす整数であり、I+1は、I<(I+1)<Mを満たす整数であり、I+2は、(I+1)<(I+2)<Mを満たす整数である\}$  に関して、次の処理を行う。

入出力コントローラ50-1は、未だ第K書き込み要求メッセージを発行していない場合、第K書き込み要求メッセージの発行を停止する。

入出力コントローラ50-1は、既に第K書き込み要求メッセージを発行し第K書き込み許可メッセージを受け取っている場合、許可フラグ76-1の該当するエントリの値を“0”に更新して、第K開放メッセージを前記第Kデータのホームであるプロセッサノード10-k ( $k=1、2、…、m$ ) のメモリコントローラ40-kにネットワーク2を介して出力する。

入出力コントローラ50-1は、既に第K書き込み要求メッセージを発行し未だ第K書き込み許可メッセージを受け取っていない場合、第K書き込み許可メッセージを受け取ったときに、許可フラグ76-1の該当するエントリの値を“1”に更新せず、第K開放メッセージの発行を行う。

第K開放メッセージを受け取ったメモリコントローラ40-kは、そのディレクトリ20-kが保持するデータの状態情報をライトロック状態情報からフリー状態情報に更新する。即ち、ディレクトリ20-2の該ブロックの値を“W、000”から“U、000”に更新する。

既に第K書き込み要求メッセージを発行したがまだ第K許可メッセージを受け取っていない場合、受け取るメッセージによって次のように動作する。第K書き込み許可メッセージを受け取った場合、許可フラグ76-1の該当するエントリの値は“0”のままとし、第K開放メッセージをメモリコントローラ40に発行する。第K不許可メッセージあるいは第K開放要求メッセージを受け取った場合はなにもしない。

【0087】

また、入出力コントローラ50-1は、開放処理フラグ79-1の値が“1”の間、第Iライトランザクションに先行する第1～第(I-1)ライトランザクションに関しては、実施例1と同じように動作する。ただし、先行する第Lライトランザクション  $\{Lは、1\leq L\leq (I-1)を満たす整数\}$  で開放要求メッセージを受け取った場合、第L書き込み要求メッセージをメモリコントローラ40-1に宛ててネットワーク2に出力する。開放処理ポインタ78-1の値を第L番目のライトランザクションであることを示す情報に設定する。そして、第Lライトを基準に、先行するライトランザクションであるか後続するライトランザクションであるかを判断し、上記開放処理を行う。

開放処理ポインタ78-1の値が示すライトランザクション（ここでは第Iライトランザクションとする）に関して第I書き込み許可メッセージを受け取ると、第I更新メッセージをメモリコントローラ40-iに発行する。また、開放処理フラグ79-1の値を“0”に更新し、開放処理を終了する。

【0088】

以上のように動作することで、例えば入出力デバイス60-1-1がライトAメッセー

ジ、ライトCメッセージの順に発行し、別の入出力デバイス60-2-2がライトCメッセージ、ライトAメッセージの順に発行した場合のデッドロックの危険性を回避することができる。

【実施例3】

【0089】

本発明のマルチプロセッサシステムは、以下に示すような動作を実行することもできる。

ここで、ディレクトリ20-iに格納される状態情報に、さらにある特定のメッセージの処理のみを受け付ける割り込み可ライトロック状態情報（以降Wiと記す）が存在する。また、割り込み可ライトロック状態時にどの入出力ノード3-jが発行したメッセージによって遷移したかを示すフィールドが加わる。以降例えば“Wi、000、2”とした場合、入出力ノード3-2が発行したメッセージによってWi状態に遷移したことを示す。

基本動作は実施例1と同様であるので、ここでは異なる部分のみを説明し、同じ部分は省略する。また、実施の最良の形態と同様に、入出力ノード3-1～3-nのうちの第1入出力ノード（入出力ノード3-1とする）の入出力デバイス60-1-1、60-1-2によって1番目からM番目（Mは1以上の整数）までのM個のデータに対するM個のライトメッセージが発行されたものとする。以降、第Iライトメッセージを例にとり説明する。

【0090】

入出力コントローラ50-1が第Iライトメッセージを受けたとき、第1から第（I-1）書き込み許可メッセージまでの（I-1）個の書き込み許可メッセージを既に受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行する。

【0091】

メモリコントローラ40-iは、第Iデータに対してフリー状態情報が前記ディレクトリ20-iに格納されているときに、第I強書き込み要求メッセージを前記入出力コントローラ50-1から受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の強・弱書き込み要求メッセージ両方を受け付けることができないライトロック状態情報を前記ディレクトリ20-iに格納する。そして、入出力コントローラ50-1に宛てて第I書き込み許可メッセージをネットワーク2に出力する。

メモリコントローラ40-iは、第Iデータに対してフリー状態情報が前記ディレクトリ20-iに格納されているときに、第I弱書き込み要求メッセージを前記入出力コントローラ50-1から受け取った場合、前記フリー状態情報に代えて、プロセッサや入出力デバイスからの読み出し要求や、前記第Iデータに対する他の弱書き込み要求メッセージを受け付けることができない割り込み可ライトロック状態情報を前記ディレクトリ20-iに格納する。そして、入出力コントローラ50-1に宛てて第I書き込み許可メッセージをネットワーク2に出力する。

メモリコントローラ40-iは、第Iデータに対してライトロック状態情報が前記ディレクトリ20-iに格納されているときに、第I強書き込み要求メッセージを前記入出力コントローラ50-1から受け取った場合、入出力コントローラ50-1に宛てて第I不許可メッセージをネットワーク2に出力する。

メモリコントローラ40-iは、第Iデータに対して割り込み可ライトロック状態情報あるいはライトロック状態情報が前記ディレクトリ20-iに格納されているときに、第I弱書き込み要求メッセージを前記入出力コントローラ50-1から受け取った場合、入出力コントローラ50-1に宛てて第I不許可メッセージをネットワーク2に出力する。

【0092】

入出力コントローラ50-1は、この第I不許可メッセージを受け取ると、第1から第（I-1）書き込み許可メッセージまでの（I-1）個の書き込み許可メッセージを既に



受け取っているか否かを検査し、受け取り済みであれば第I強書き込み要求メッセージを、受け取り済みでなければ第I弱書き込み要求メッセージを発行する。

メモリコントローラ40-iは、第Iデータに対して割り込み可ライトロック状態情報が前記ディレクトリ20-iに格納されているときに、第I強書き込み要求メッセージを前記入出力コントローラ50-1から受け取った場合、前記割り込み可ライトロック状態情報に代えて、リクエストロック状態情報を前記ディレクトリ20-iに格納する。そして、前記ディレクトリに格納された入出力コントローラを特定する情報（ここでは入出力コントローラ50-2とする）が指す入出力コントローラ50-2に宛てて第I再試行要求メッセージをネットワークに2出力する。

第I再試行要求メッセージを受け取った入出力コントローラ50-2は、以下に示す再試行処理を行う。

【0093】

まず、メモリコントローラ40-iに宛てて第I開放メッセージをネットワーク2に出力する。

次に、入出力コントローラ50-2が受け取ったライトメッセージの中で、前記第Iデータとアドレスを同じくするライトメッセージで書き込み許可メッセージを受け取り済みのものに関して、まだ更新メッセージを発行していなければ更新メッセージの発行を停止する。そして、前記第I開放メッセージの発行後に、メモリコントローラ40-iに宛てて書き込み要求メッセージを発行する。これで再試行処理は終わる。

メモリコントローラ40-iは、前記第I開放メッセージを受け取ると、リクエストロック状態情報に代えてライトロック状態情報を前記ディレクトリ20-iに格納する。そして、入出力コントローラ50-1に宛てて第I書き込み許可メッセージを前記ネットワークに出力する。

また、メモリコントローラ40-iは、第Iデータに対してリクエストロック状態情報が前記ディレクトリ20-iに格納されているときに、第Iデータに対する更新メッセージを前記入出力コントローラ50-2から受け取った場合、主記憶部30-2のデータの値を更新メッセージで指定される値に更新する。

【0094】

また、第3実施例の変形として、次のような変形も可能である。

第I再試行要求メッセージを受け取った入出力コントローラ50-2が行う再試行処理で、まず、入出力コントローラ50-2が受け取ったライトメッセージの中で、前記第Iデータとアドレスを同じくするライトメッセージで書き込み許可メッセージを受け取り済みのものに関して、まだ更新メッセージを発行していなければ第I開放メッセージを発行し、その後にメモリコントローラ40-iに宛てて書き込み要求メッセージを発行し処理を終える。また、まだ更新メッセージを発行していればなにもせず処理を終える。

メモリコントローラ40-iは、前記第I開放メッセージを受け取ると、リクエストロック状態情報に代えてライトロック状態情報を前記ディレクトリ20-iに格納する。そして、入出力コントローラ50-1に宛てて第I書き込み許可メッセージを前記ネットワークに出力する。

また、メモリコントローラ40-iは、第Iデータに対してリクエストロック状態情報が前記ディレクトリ20-iに格納されているときに、第Iデータに対する更新メッセージを前記入出力コントローラ50-2から受け取った場合、主記憶部30-2のデータの値を更新メッセージで指定される値に更新する。そして、リクエストロック状態情報に代えてライトロック状態情報を前記ディレクトリ20-iに格納する。そして、入出力コントローラ50-1に宛てて第I書き込み許可メッセージを前記ネットワークに出力する。

【0095】

以上のように構成、動作することで、実施例2と同じようにデッドロックの危険性を回避することができる。

【図面の簡単な説明】

【0096】

【図1】従来のマルチプロセッサシステムの構成を示す図である。

【図2】従来技術1の動作を示すタイミングチャート図である。

【図3】従来技術2の動作を示すタイミングチャート図である。

【図4】本発明のマルチプロセッサシステムの構成を示す図である。

【図5】本発明のマルチプロセッサシステムにおける入出力コントローラの構成を示す図である。

【図6】本発明のマルチプロセッサシステムの動作として、ライトメッセージが連続して発行された場合の動作を示すタイミングチャート図である。

【図7A】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7B】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7C】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7D】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7E】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7F】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7G】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7H】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図7I】本発明のマルチプロセッサシステムにおける入出力コントローラの動作例を説明するための図である。

【図8】本発明のマルチプロセッサシステムの他の構成を示す図である。

【図9】本発明の実施例1に係るマルチプロセッサシステムの動作として、ディレクタの状態がUncachedであった場合の動作を示すタイミングチャート図である。

【図10】本発明の実施例1に係るマルチプロセッサシステムの動作として、ディレクタの状態がCleanであった場合の動作を示すタイミングチャート図である。

【図11】本発明の実施例1に係るマルチプロセッサシステムの動作として、ディレクタの状態がDirtyであった場合の動作を示すタイミングチャート図である。

【図12】本発明の実施例1に係るマルチプロセッサシステムの動作として、ディレクタの状態がリクエストロック状態あるいはライトロック状態であった場合の動作を示すタイミングチャート図である。

【図13】本発明の実施例2に係るマルチプロセッサシステムにおける入出力コントローラの構成を示す図である。

【図14】本発明のマルチプロセッサシステムにおける入出力コントローラが、第I書き込み許可メッセージを受けた場合の動作を示すフローチャート図である。

【符号の説明】

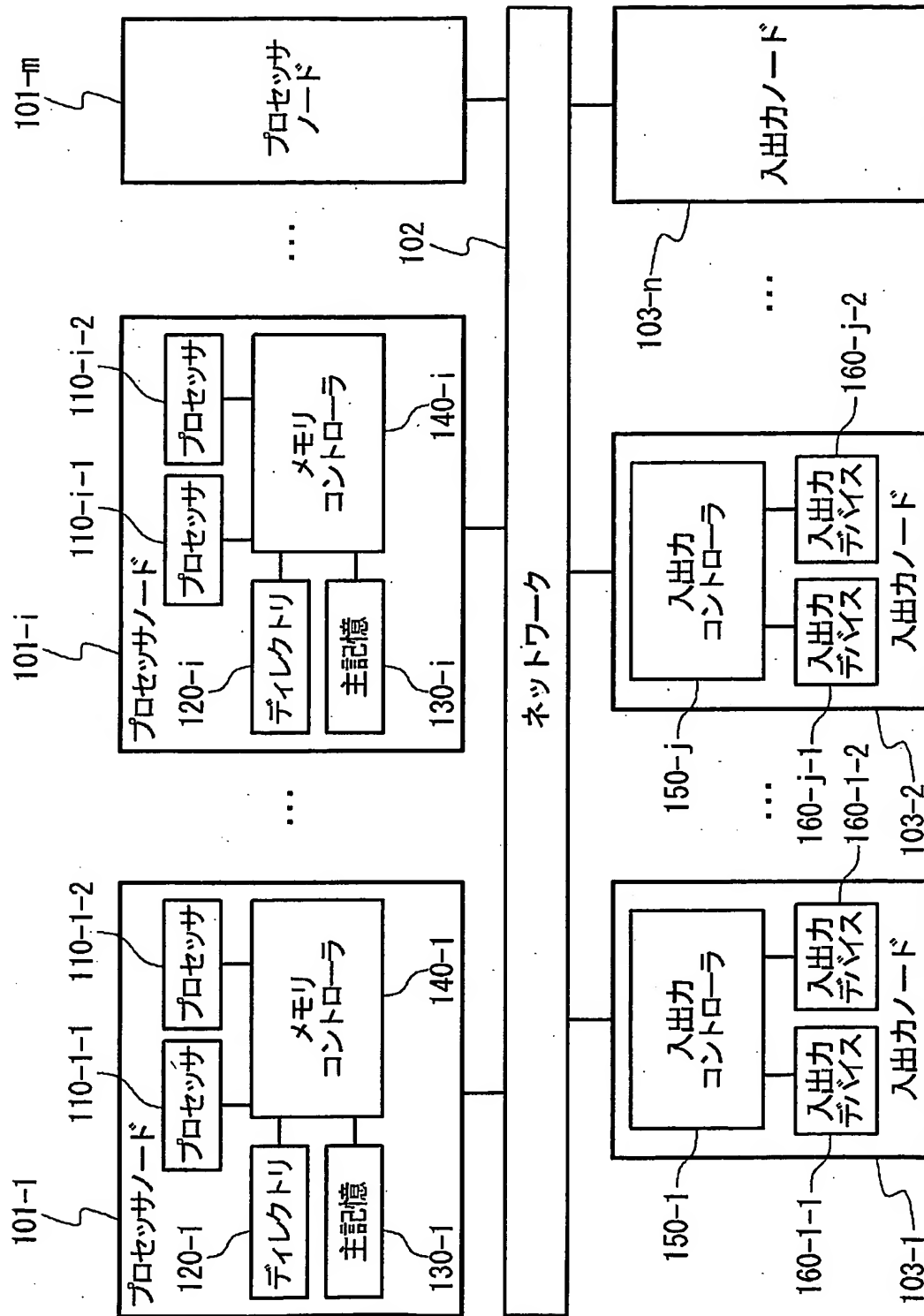
【0097】

- 1-1~1-m      プロセッサノード
- 2      ネットワーク
- 3-1~3-n      入出力ノード
- 10-i-1、10-i-2 (i=1、2、…、m)      プロセッサ
- 20-i (i=1、2、…、m)      ディレクトリ
- 30-i (i=1、2、…、m)      主記憶
- 40-i (i=1、2、…、m)      メモリコントローラ

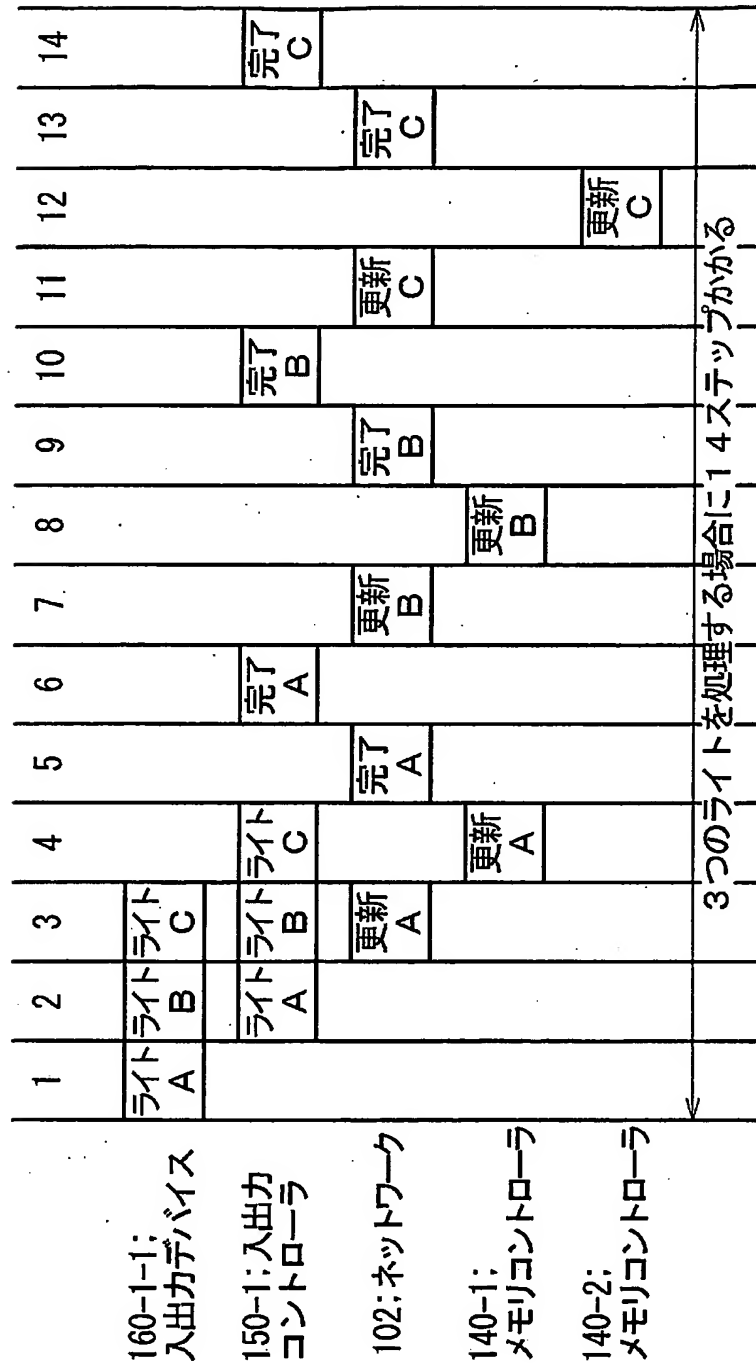
50-j (j=1, 2, ..., n) 入出力コントローラ  
51-j (j=1, 2, ..., n) 入出力セレクタ  
52-j (j=1, 2, ..., n) 入出力コントローラ  
60-j-1, 60-j-2 (j=1, 2, ..., n) 入出力デバイス  
71-j (j=1, 2, ..., n) セレクタ  
72-j (j=1, 2, ..., n) メッセージ格納キュー  
73-j (j=1, 2, ..., n) ライトポインタ  
74-j (j=1, 2, ..., n) リードポインタA  
75-j (j=1, 2, ..., n) リードポインタB  
76-j (j=1, 2, ..., n) 許可フラグ  
78-j (j=1, 2, ..., n) 開放処理ポインタ  
79-j (j=1, 2, ..., n) 開放処理フラグ  
101-1~101-m プロセッサノード  
102 ネットワーク  
103-1~103-n 入出力ノード  
110-i-1, 110-i-2 (i=1, 2, ..., m) プロセッサ  
120-i (i=1, 2, ..., m) ディレクトリ  
130-i (i=1, 2, ..., m) 主記憶  
140-i (i=1, 2, ..., m) メモリコントローラ  
150-j (j=1, 2, ..., n) 入出力コントローラ  
160-j-1, 60-j-2 (j=1, 2, ..., n) 入出力デバイス

【書類名】 図面

【図 1】



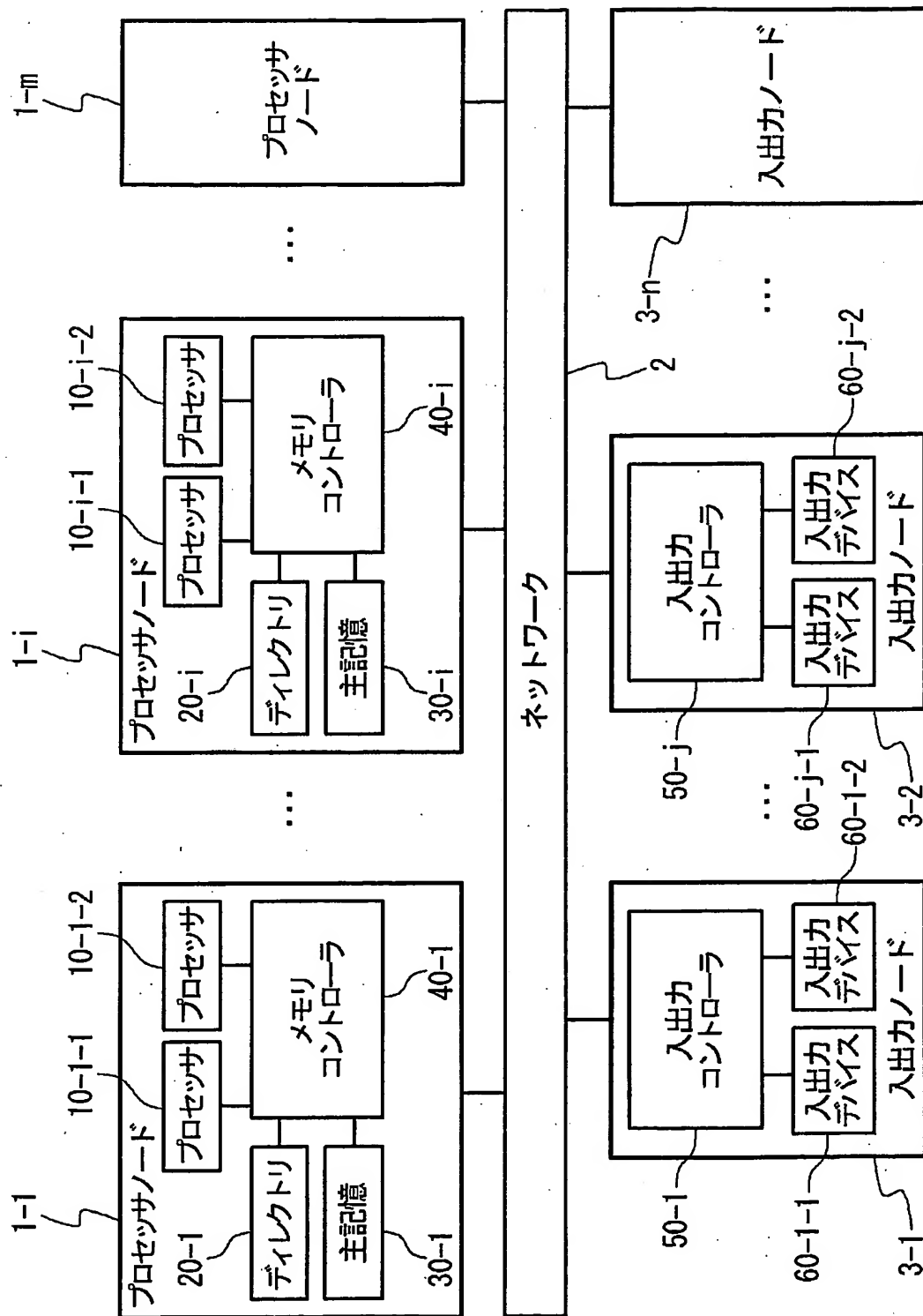
【図2】



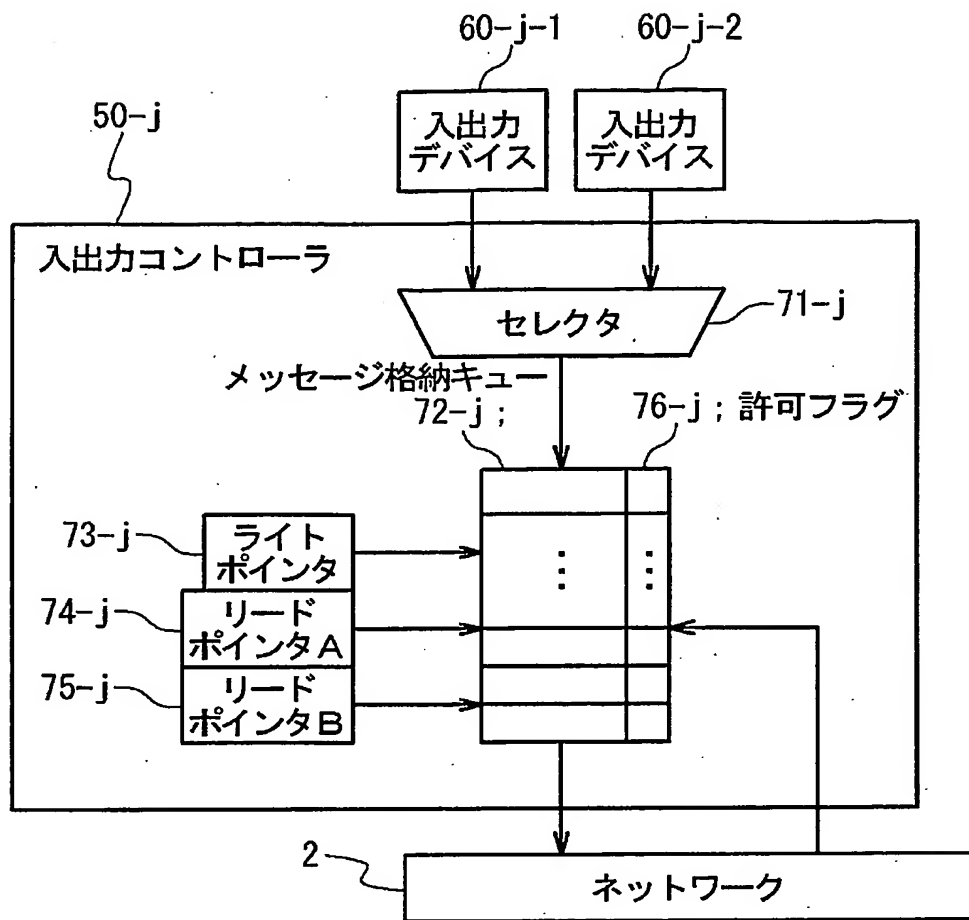
【図 3】

	1	2	3	4	5	6	7	8	9	10	11
160-1-1: 入出力 デバイス	ライト A	ライト B	ライト C								
150-1:入出力 コントローラ		ライト A	ライト B	ライト C		完了 A	完了 B				完了 C
102: ネットワーク			更新 A	更新 B	完了 A	完了 B		更新 C		完了 C	
140-1:メモリ コントローラ				更新 A	更新 B						
140-2:メモリ コントローラ									更新 C		
	← 3つのライトを処理する場合に1.1ステップかかる →										

【図 4】



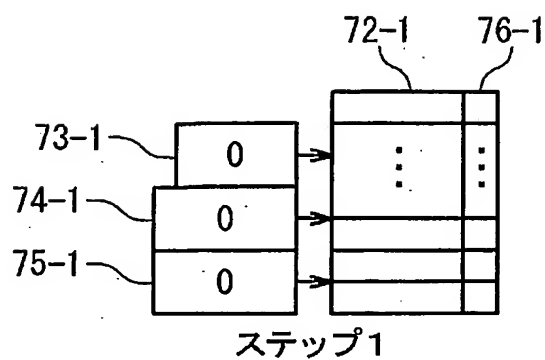
【図 5】



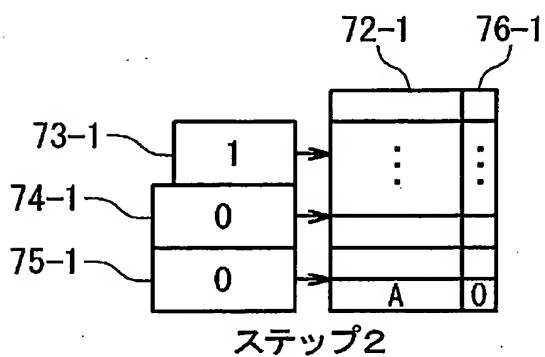




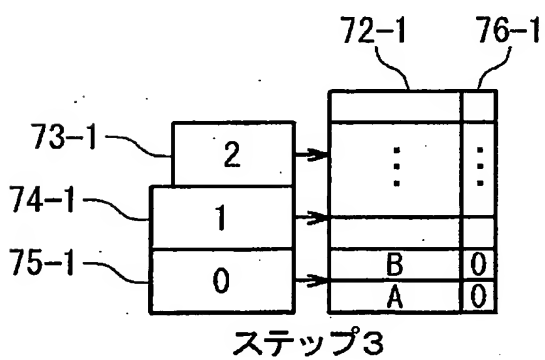
【図 7 A】



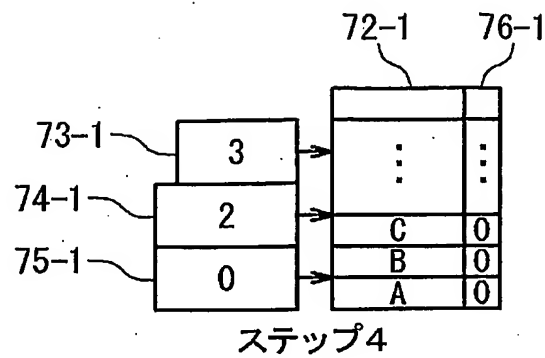
【図 7 B】



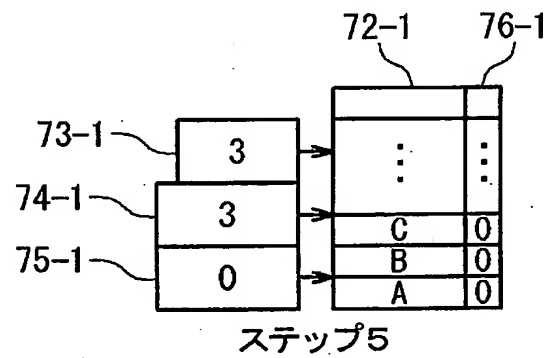
【図 7 C】



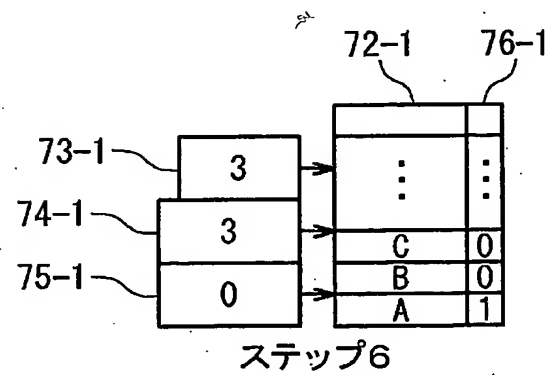
【図 7 D】



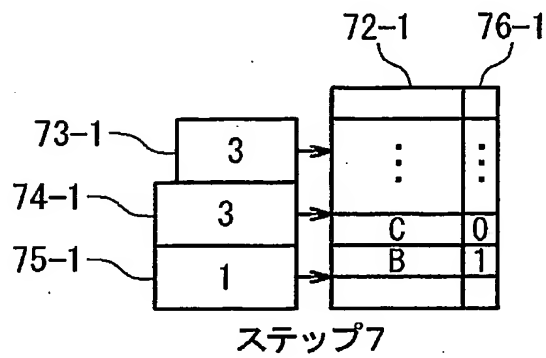
【図 7 E】



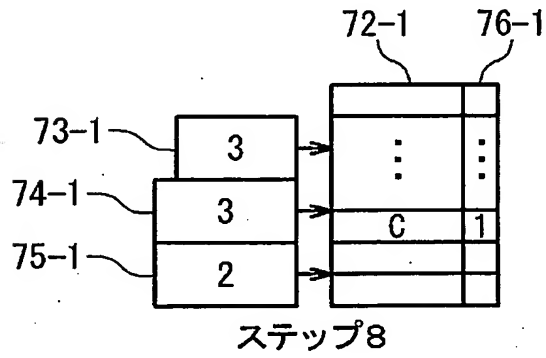
【図 7 F】



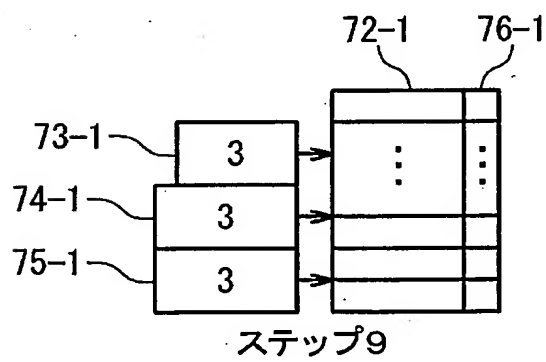
【図 7 G】



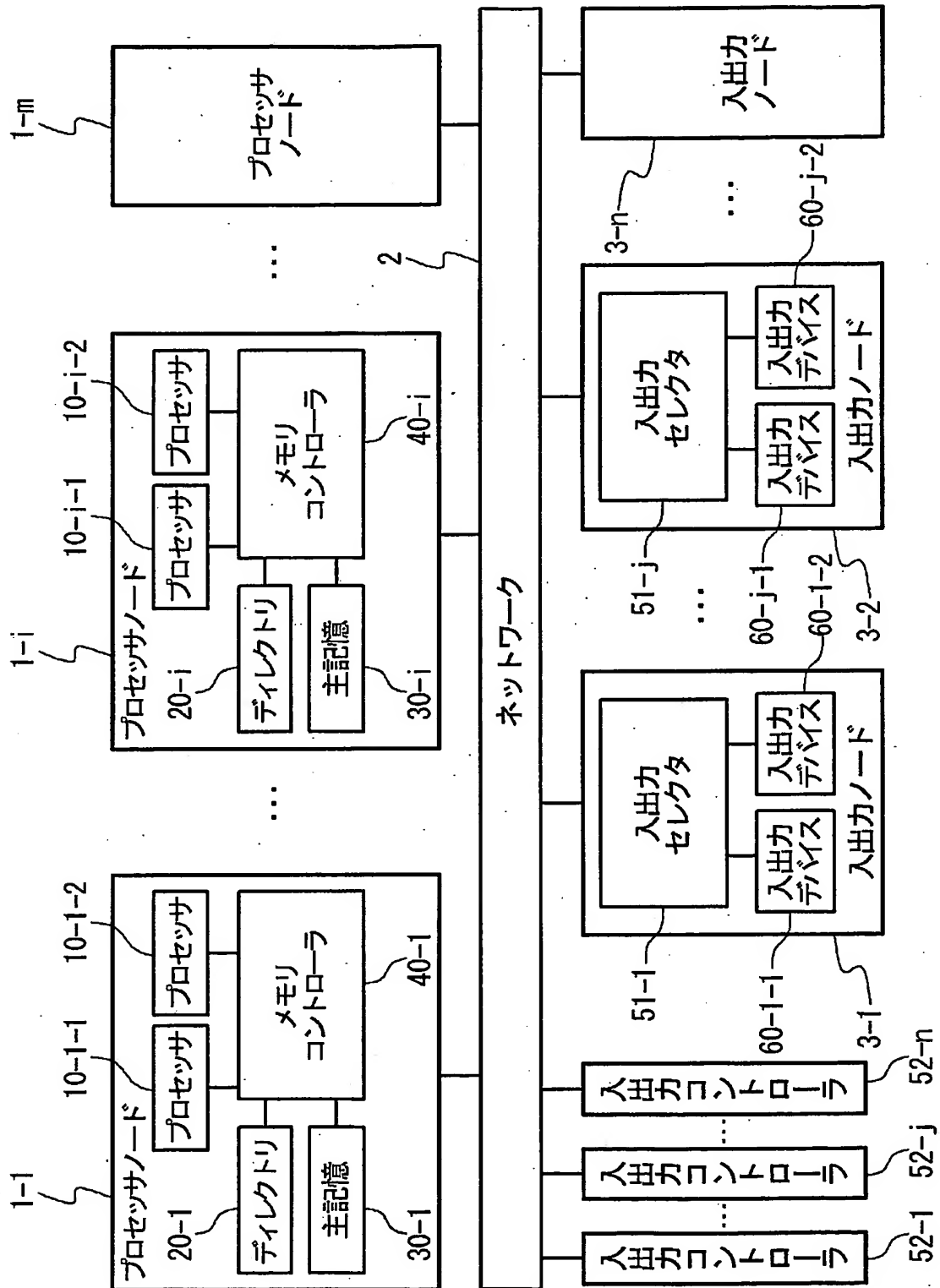
【図 7 H】



【図 7 I】



【図 8】



【図 9】

	1	2	3	4	5	6	7	8	9
60-1-1; 入出力デバイス	ライト A								
50-1; 入出力コントローラ		ライト A				許可 A			
2; ネットワーク			要求 A		許可 A		更新 A		
40-1; メモリコントローラ				要求 A				更新 A	
20-1; ディレクトリ	U, 000			W, 000				U, 000	



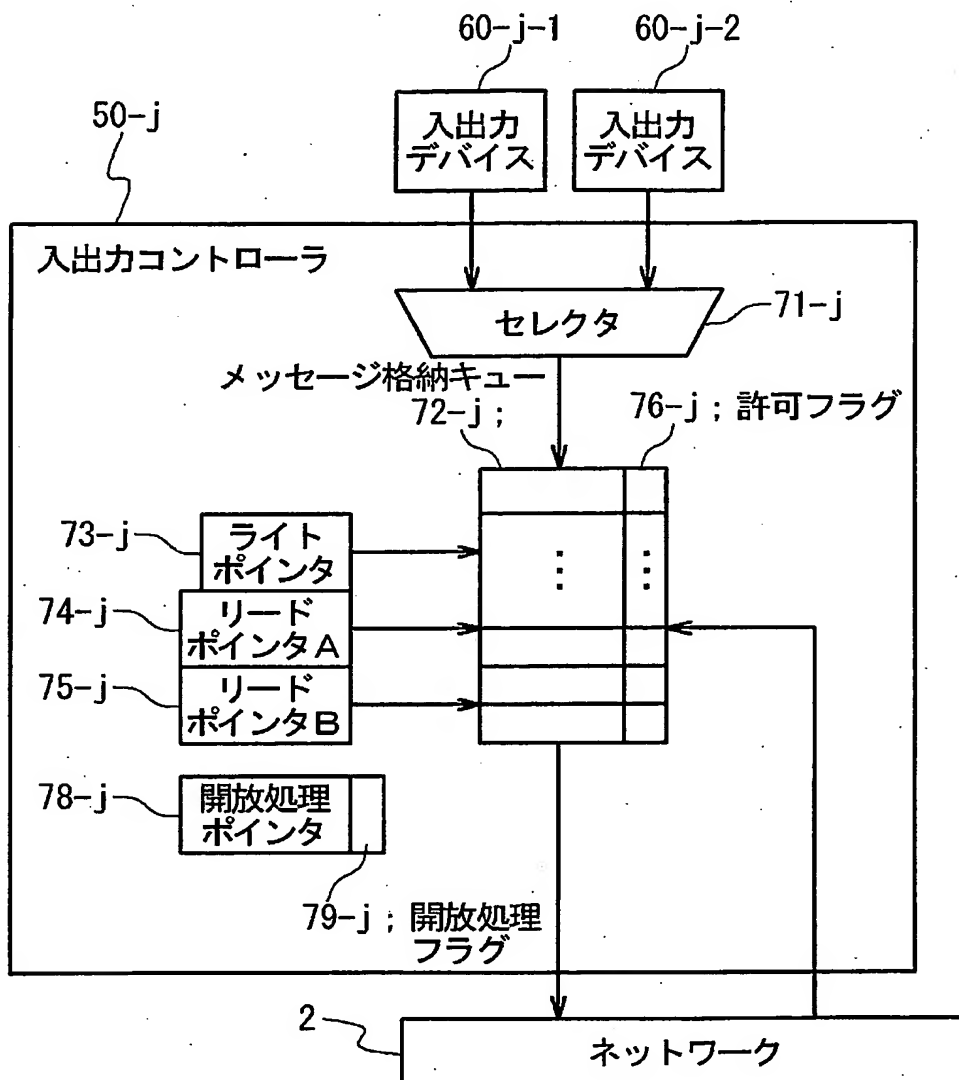




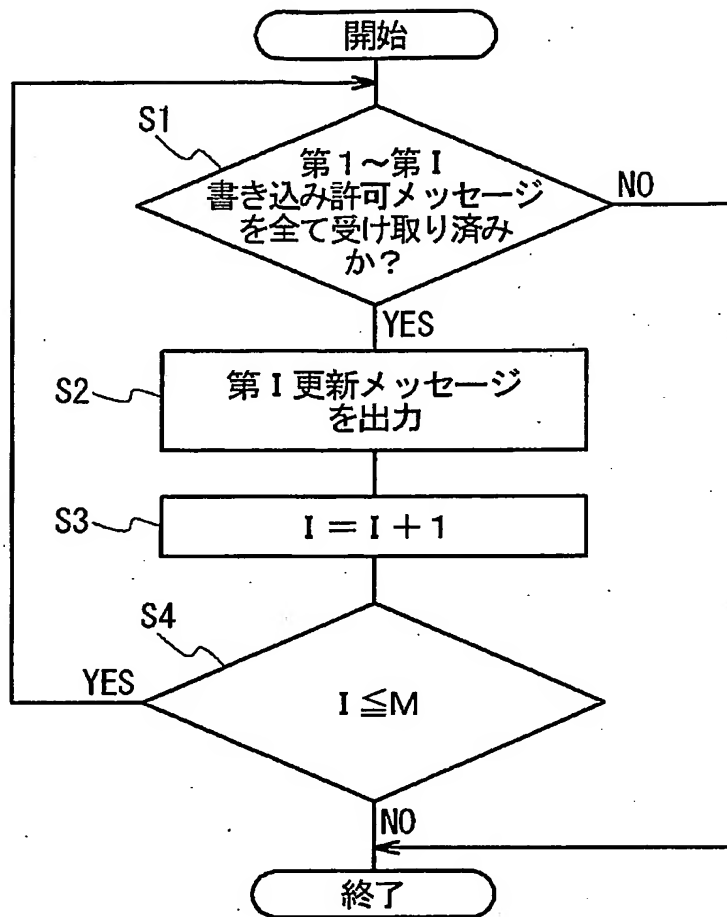
【図 12】

	1	2	3	4	5	6	7
60-1-1; 入出力デバイス	ライト A						
50-1; 入出力コントローラ		ライト A				不許可 A	
2; ネットワーク			要求 A		不許可 A		要求 A
40-1; メモリコントローラ				要求 A			
20-1; ディレクトリ	R, --- or W, 000						

【図 13】



【図 14】



**【書類名】 要約書****【要約】**

**【課題】** 入出力コントローラが、異なるプロセッサノードを宛先とする入出力デバイスからの複数のライトメッセージを連続して処理することができるマルチプロセッサシステムを提供すること。

**【解決手段】** 入出力コントローラがライトメッセージを受けたときに、該当するデータをメモリに保持するホムプロセッサノードに書き込み要求メッセージを発行する。書き込み要求メッセージを受け取ったプロセッサノードのメモリコントローラは、ディレクトリに格納された該当するデータの状態に基づいて一貫性処理を行い、該書き込み要求メッセージを発行した入出力コントローラに書き込みの許可を示すメッセージが届くように制御する。書き込みの許可を表すメッセージを受け取った入出力ノードの入出力コントローラは、データの書き込みを行うライトメッセージとして更新メッセージをホムのプロセッサノードに発行する。更新メッセージを受け取ったプロセッサノードのメモリコントローラは、主記憶部のデータを更新する。上記処理で、入出力コントローラは、入出力デバイスから複数のライトメッセージを受け取ったときに、先行するライトメッセージの進捗にかかわらず書き込み要求メッセージを発行し、先行するライトのライトメッセージ発行が行われた後にライトメッセージを発行する。

**【選択図】** 図 4

特願2004-197296

出願人履歴情報

識別番号

[000004237]

1. 変更年月日

1990年 8月29日

[変更理由]

新規登録

住所

東京都港区芝五丁目7番1号

氏名

日本電気株式会社

特願2004-197296

出願人履歴情報

識別番号

[000168285]

1. 変更年月日  
[変更理由]

2002年 7月30日

名称変更

住所変更

住 所  
氏 名

山梨県甲府市大津町1088-3

エヌイーシーコンピュータテクノ株式会社